NATIONAL INSTITUTES OF HEALTH

National Library of Medicine

Programs and Services

Fiscal Year 2010

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

National Library of Medicine Catalog in Publication

National Library of Medicine (US)
National Library of Medicine programs and services.—
Bethesda, Md.: The Library, [1978-]
1977 –
v.:
Report covers fiscal year.
Continues: National Library of Medicine (US). Programs and Services.
Vols. For 1977-78 issued as DHEW publication; no. (NIH) 78-256, etc.;
for 1979-80 as NIH publication; no. 80-256, etc.
Vols. 1981 – present Available from the National Technical Information Service,
Springfield, Va.
Reports for 1997 – present issued also online.
ISSN 0163-4569 = National Library of Medicine programs and services.

1. Information Services – United States – Periodicals 2. Libraries, Medical –
United States – Periodicals I. Title II. Title: National Library of Medicine
programs & services III. Series: DHEW publication; no. 78-256, etc. IV.
Series: NIH publication; no. 80-256, etc.

Z 675.M4U56an

CONTENTS

Tables

PREFACE

I'm pleased to present this report on the programs and services of the National Library of Medicine. Although it houses the world's largest biomedical collection, NLM is about much more than books, journals, artworks and audiovisuals. Each day, via the Internet, NLM delivers trillions of bytes of health data crucial to the lives of millions of people around the globe. The Library also conducts and funds research and development across a wide range of disciplines, all with the aim of improving the public health. Notable achievements in Fiscal Year 2010 include the following:

- ClinicalTrials.gov celebrated a decade of public service. This important resource now contains data on over 97,000 federally and privately supported studies, including summary results, adverse event information and other features.

- PubMed Central (PMC), the free online repository of peer-reviewed biomedical research reports, added its 2 millionth article on July 27, 2010. That same day, PubMed, the world's largest biomedical database, added its 20 millionth citation. PMC observed its tenth anniversary in 2010.

- Interoperability and standardization of terminology are key elements of the national electronic health record effort. NLM has a long and distinguished history in this area and has worked closely with policymakers at HHS and the White House, and continued its own research.

- This year, the Library added resources to the disaster information management Web site, to assist those coping with earthquakes in Chile and Haiti, and the Gulf oil spill. NLM also activated the Emergency Access Initiative, to provide free reference materials to Haiti and flood-ravaged Pakistan.

- 2010 saw the second year of American Recovery and Reinvestment Act funding. Overall, NLM invested nearly $84 million in basic and applied research projects, expanding research and speeding discovery in the name of improving public health.

- MedlinePlus, the popular consumer health Web site, got a well-received makeover. The print and online versions of the *NIH MedlinePlus* magazine and its Spanish-English counterpart, *Salud*, were distributed to hundreds of thousands of outlets nationwide.

- Social media really took off at NLM this year. Facebook, Twitter and YouTube helped NLM spread the word about its many programs and services to audiences around the world.

These are just some of NLM's significant contributions to its mission to acquire, organize and disseminate biomedical information for the betterment of American–and global–human health. I commend the staff and our many advisors and consultants for making the National Library of Medicine an institution revered around the world. Thank you.

Donald A.B. Lindberg, MD
Director

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Milton Corn, MD
Acting Associate Director

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

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

Planning and Analysis

The NLM Long Range Plan remains at the heart of NLM's planning and budget activities. Its goals form the basis for NLM operating budgets each year. *Charting a Course for the 21st Century: NLM's Long Range Plan 2006–2016* is available in print and on the NLM Web site. Print copies are available from the NLM Office of Communications and Public Liaison. The report 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; and
- Goal 4. A Strong and Diverse Workforce for Biomedical Informatics Research, Systems Development, and Innovative Service Delivery.

Outreach and Consumer Health

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

In addition to specific outreach and consumer health projects outlined below, OHIPD has overall responsibility for developing and coordinating the NLM Health Disparities Plan. This plan outlines NLM strategies and activities undertaken in support of NIH efforts to understand and eliminate health disparities between minority and majority populations. NLM's Health Disparities Plan is available on the NLM Web site.

The Information Prescription Project ("Information Rx"), initiated with the American College of Physicians in 2003, continues to grow. Following collaboration in prior years with chapters of the National Medical Association and Student National Medical Association to introduce the Information Rx concept to their membership, a project with the 43,000-member American Academy of Physician Assistants (AAPA) was conducted in FY2010, in collaboration with OCPL and the NN/LM. Other organizations involved in the Information Rx project include the American Osteopathic Association, hospital librarian members of the Medical Library Association, and disease-focused organizations such as the Fisher Center for Alzheimer's Research.

OHIPD staff are also pursuing outreach initiatives intended to encourage underrepresented minority high school students to pursue careers in medicine and the health sciences, carried out in collaboration with other divisions of NLM.

An initiative with ExploreHealthCareers.com and Mentoring In Medicine (MIM), which began in FY2008, continues to support an annual workshop entitled "Yes, I can be a Healthcare Professional." This is an inspiring and well-attended workshop convened for parents and children in grades three through college. This program, which pairs students with more than 500 healthcare 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. A new MIM Program that was started in FY2009, Science and Health Career Exploration, continues to be funded by NLM and co-sponsored by the Friends of the NLM. This initiative is designed to reach six public and charter schools with an after school program to enrich the high school biology curriculum and encourage enrollment in higher education programs leading to degrees in medicine,

allied health professions, and medical librarianship. Principals, science teachers and guidance counselors from participating schools oversee 40 sessions of biology instruction in 12 organ systems, taught by visiting health professionals/mentors over a two-year period. The participating schools constitute an ideal laboratory in which to experiment with new approaches for stimulating and sustaining student interest in health careers, within the larger context of STEM education in the US. During its two-year operation, the program has exposed over 600 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. During the Fall 2010 semester, the evaluation findings revealed an impressive gain 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.

A collaboration with the Student National Medical Association supports that organization's efforts to encourage African-American medical students to pursue careers as physician researchers. NLM's support is focused on promoting consideration of research careers in biomedical and public health informatics. A pipeline for a strong and diverse workforce is a high priority goal.

The Lower Rio Grande Valley Hispanic Outreach Project was a collaboration with the University of Texas at San Antonio Health Sciences Center to conduct a needs assessment and various health information outreach projects with Hispanic-serving community, health, and educational institutions. This was the beginning of an intensified NLM effort to meet the health information needs of the Hispanic population in Texas and elsewhere. In FY2010, NLM continued its support for the South Texas High School for the Health Professions, known as MedHigh, a magnet health high school in the Lower Rio Grande Valley of Texas. The MedHigh VIVA! Peer Tutors Program is an award winning effort to involve high schools students in teaching their peers about online health information. The peer tutors also conduct outreach to the local community and sponsor annual online virtual conferences open to interested faculty, librarians, and students from high schools around the country. MedlinePlus en español is being emphasized where applicable. The project was expanded to include a health careers tutoring component that involves the students, teachers, and guidance counselors. 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.

Native American Outreach

In 2010, 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 11 pow-wows, mostly in the Mid-Atlantic area. An estimated 5,800 persons visited the NLM booth over the course of these pow-wows. These activities proved to be another viable way to bring NLM's health information to the attention to segments of the Native American community and the general public. OHIPD also supported two projects in the Dakotas that resulted largely from the Native American Listening Circles conducted in prior years and have continued into FY2010.

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 outreach activities run by the library.

At MHA Systems Inc., a tribal enterprise of the MHA Nation, a continuing economic development outreach project provides outreach assistance to a tribal information technology company that would ultimately result in jobs creation on the reservation (in this case, the Ft. Berthold Indian Reservation). The project is intended to improve the competitive capabilities of MHA Systems Inc., and also to refine, test, and strengthen the company's core scanning services. Current projects include converting thousands of abstracts from AIDS-related conferences from paper and CDROM format into NLM-approved XML electronic formats, and transcribing taped interviews related to the planned NLM exhibition on "Native Concepts of Health and Illness."

Tribal Consultations on Exhibition Concepts

In FY2010, OHIPD participated with other NLM divisions in coordinating and facilitating a consultation activity with the Mississippi Band of Choctaw Indians in Choctaw, MS. OHIPD arranged for the NLM Director to conduct interviews with physicians, traditional healers, and community leaders on topics relevant to the planned NLM exhibition on "Native Concepts of Health and Illness." This was the fifth such consultation activity, with four others having been conducted in prior years in Anchorage, AK, Santa Fe, NM, Seattle, WA, and the Hawaiian Islands.

OHIPD organized a workgroup meeting with representation from varied tribes, convened on September 29-30, 2010, at NLM. The purpose of the meeting was to obtain advice on the scope, content, and implementation of the exhibition. The meeting included American Indians, Alaska Natives, and Native Hawaiians drawn from tribal and Western medical, health, and cultural leaders. Participants discussed a variety of ideas, topics, and perspectives for possible use in the exhibition.

Evaluation

The Internet and World Wide Web play a dominant role in dissemination of NLM information services, and the Web environment in which NLM operates is rapidly changing and intensely competitive. These two factors combined suggested the need for comprehensive and dynamic NLM Web planning and evaluation process. The Web evaluation priorities of the OCHD include both quantitative and qualitative metrics of Web usage and measures of customer perception and use of NLM Web sites. During FY2010, 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; and access to Internet audience measurement estimates based on Web usage by user panels organized by a private sector company. During FY2010, OHIPD continued to coordinate NLM's use of the online Web user survey known as the American Customer Satisfaction Index (ACSI). The ACSI provides ongoing user feedback to NLM's Web site manager.

International Programs

NLM's international partnerships and projects strengthen and expand global access to the world's health literature. The focus of the Office of International Programs is on outreach to researchers, physicians, and librarians in developing countries, with an additional, more recent emphasis on medical students, health workers and end users. This office continues to develop pilot programs, dissemination strategies, and training opportunities as well as evaluation, presentation, and publication of results.

Originally targeting one area of opportunity in 1997, NLM played a critical role in the Multilateral Initiative on Malaria. Since that time, NLM's international outreach efforts have focused in particular on Africa, where the need is great. All programs are based on making information available to and integrating information from areas where disease is endemic. Building on 13 years of NLM engagement in sub-Saharan Africa, these programs also often support, connect with, and sustain one another. For example, librarians are beginning to play a more substantial role in the project with medical journal editors. One librarian served the malaria researchers with full-text articles, prior to WHO's HINARI, in the early days of the malaria research network, and the medical students are being trained by librarians in searching NLM databases, rather than using Google and Wikipedia as first choices.

NLM has applied greater focus on global health by piloting demonstration projects that draw strength from one another and tie into NLM's major programs and databases. These areas of emphasis touch and strengthen all phases of the research process—from journal editors and librarians to today's scientists and those of the future.

The African Medical Journal Editors Partnership Program strengthens six African journals for acceptance into MEDLINE through capacity building and partnerships with six major medical journals in the US and UK; this program makes important research being carried out in endemic countries available to the world. The NLM Associates Program has accepted six fellows from the African continent over the past 10 years; they now compose a network of former African associates and are having a demonstrable impact at seven universities in seven African countries. NLM has guided African medical students in the creation of MedlinePlus African tutorials on malaria and diarrhea. Implementing these tutorials as part of an "information intervention" at the village level has resulted in small research projects and capacity building for the future. These programs all build on NLM's leadership in the Multilateral Initiative on Malaria program, in which 27 research sites in 14 African countries were connected to the Internet and to medical literature.

Information Communication Technology (ICT)

NLM's initial leadership for the Multilateral Initiative on Malaria (http://www.mimalaria.org/) more than ten years ago has led to more recent support in FY2010 for the implementation of an electronic health management information system (eHMIS - http://www.ehmis.net/) at Tororo Hospital in Eastern Uganda, the first in the country. The first phase of this project was the implementation of the system in the hospital and training of staff, followed by the second phase of that connects a health center to the system and makes medical information available at point of care.

Also in FY2010, NLM initiated a community-based research project to utilize digital pen technology, originally developed for an NLM disaster response application, as a data capture method in an observational study on actual bednet use in Africa. The study was carried out in Mifumi village in the Eastern Uganda region, a village that had also received the MedlinePlus tutorial on malaria in prior years, by a team of Makerere University medical students. The students were supported by two technical teams, one at NLM and one in Uganda, as well as the nurse sister of the health center and the Mifumi village community. Their quest was to see whether the 300 bednets distributed in 2006 were actually being used to prevent malaria and to evaluate how digital pen technology could support efficient and timely collection, collation, and presentation of the research data regarding use of the nets. One researcher reported it to be "a good opportunity to use technology for capturing raw data in the community in a fast and efficient manner." Another said it was "really good experience to capture research information which is then instantly transferred to a database ready for analysis. This reduces potential data entry errors."

Although malaria prevalence remains high, the trend in morbidity and mortality in the village since a collaborative information intervention with MedlinePlus

tutorials that began three years ago is downward. The study results showed that, generally, households endeavor to use mosquito nets, but many were damaged, not being used in the most effective manner, or used only on some family members. A community meeting in the village to present results of the survey to the people in village was attended by over 150 people and lasted for two hours. The people very much appreciated that the research was being presented to them, and that their ideas were being considered.

Network of African Librarians

NLM continues its commitment to utilizing and expanding the leadership of a growing network of African librarians who have received training as NLM Associate Fellows (http://www.nlm.nih.gov/about/training/associate/africannetwork.html) or as MLA Cunningham Fellows. The objectives of supporting this network is to assist African librarians who already have links to NLM in creating an approach for strengthening libraries through outreach and training in Africa, and to explore how this librarian corps can be brought together with the African Medical Journal Editor Partnership Project (Mali, Uganda, Malawi, Ghana, Zambia, Ethiopia) and African research and clinical communities. The network currently comprises seven librarians from Kenya, Zambia, Mozambique, Mali, Nigeria, Uganda, and Zimbabwe.

The NLM Associates program and the significant role librarians play in other NLM projects have led to creation of Network of African Medical Librarians and Deans. In response to the librarians' interest in creating a course on information retrieval in FY2001, and in collaboration with Kenyatta University, NLM convened a groundbreaking meeting of medical librarians and deans/provosts (or their representatives) from medical schools where the librarians are based. The objective was to develop a course to be adopted at their respective universities as a permanent part of the medical school curriculum. The deans welcomed the initiative, and participants engaged in a lively discussion regarding strategies for expanding the course to encompass basic computer literacy training and the writing of papers for journal publication (and thereby strengthening medical journals) and finally for integrating the course into the medical school curricula. The Web site created by NLM's International Programs office for the meeting is: http://karibouconnections.net/wordpress/medlibafrica/. Based on the enthusiasm generated at this meeting, the librarians collaborated online to produce course modules. The course was launched at the Association for Health Information and Libraries in Africa (AHILA) 2010 international congress in Ougadougou, Burkina Faso. There is already interest in use of this course outside of the original institutions involved—from the African Medical Journal Editors Partnership Project to bioinformatics workshop training in Morocco to ministry-level programs in Haiti.

At their respective libraries, these librarians continue to be busy training faculty and students as well as engaging in outreach to areas outside of the capital cities. They have carried out workshops for librarians and researchers from around their countries, produced regular newsletters, presented at faculty board meetings, and conducted lunchtime training sessions for staff. Several have developed institutional repositories which can be accessed online from anywhere.

NLM's International Progams office also collaborates with the NLM Associates Program in identifying future associates from Africa as well as in structuring the curriculum to support their work when they return to their home libraries. The 2010-11 African Associate is from the National Institute of Hygiene in Morocco.

MedlinePlus African Tutorials

This project is another effort by NLM to reach the consumer/end user, no matter where that user is located. MedlinePlus African tutorials focus on tropical disease issues in developing country contexts. The first two tutorials (http://www.nlm.nih.gov/medlineplus/africa/) were on malaria and diarrhea and were developed with the Faculty of Medicine at Makerere University in Uganda, field tested and distributed in FY2008. As of FY2010, these tutorials have been expanded to included versions in six local languages.

Another tutorial under development focuses on mental health and is led by the dean and team of students from Gulu Medical School, located in the epicenter of the Northern Uganda region which was torn by war for 20 years. The text was written by medical school students and faculty while the illustrations are being created by children from around the region, many victims of the conflict. In field testing, a pre-test survey on depression has been administered and draft tutorial booklet has been read with post-testing in several settings—NGO, cultural village, family protection unit of police, Army, detention prison, college for teachers, schools. Field testing is revealing that a number of the terms used in the tutorial are too clinical and not understood by laypeople. The final version will use simpler language. All find the information useful. After field testing has concluded, a final version of the tutorial as well as a training manual for community use will be published.

A new collaboration has continue work on community education materials for Burkitt's lymphoma study with NCI principal investigator, Ugandan team of medical staff at two hospitals, artists and medical student team. Completed materials include Burkitt's lymphoma tutorial, video, pictorial informed consent, flyer, posters of Burkitt's lymphoma awareness and why people don't seek early treatment—all reviewed by medical teams in Kampala and Gulu.

Web Sites

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

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

Malaria Research Resources (http://www.nlm.nih.gov/mimcom/mimcomhomepage.html) supports the activities of MIMCom, a project of the Multilateral Initiative on Malaria and the National Library of Medicine to support African scientists and malaria researchers in their ability to connect with one another and sources of information through full access to the Internet and the resources of the World Wide Web. Having established or enhanced connectivity at 19 research sites in 13 countries, NLM's current focus is on products and databases to aid the efforts of malaria research.

MIMCom News, a weekly newsletter started by this office, is now a privately run publication known as *MalariaWorld* (www.malariaworld.org). *Malaria World* provides the latest information on malaria every week, reaching over 6,000 subscribers. The newsletter aims to be the most complete electronic malaria information resource, covering announcements, contributions from subscribers, scientific publications, reports, events, jobs, grants, training and research opportunities, and news.

African Medical Journal Editors Partnership Program

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

With the support of the Partnership Project, staff from each African journal visited the offices of its partner journal for one to two weeks. African editors reported these site visits to be extremely useful for observing the editorial and publishing practices of another journal.

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

Trans-NIH Collaboration

International Programs also: participates in the newly formed NIH Global Health Research Working Group and two subcommittees (one to create an NIH-wide International Activity Database and another to make input to the 2011 UN General Assembly high level meeting of heads of government on non-communicable disease); the NIH mHealth Working Group and mHealth Summit; the US-Pakistan roundtable to contribute to bilateral strategic dialog meeting in Islamabad; preparation for G8 Consultation meeting held at NIH; and teaching class in Global Health for FAES.

International Programs 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.

Visitors

In FY2010, The Office of Communications and Public Liaison, and the History of Medicine Division's Exhibition Program arranged 213 tours—115 regular, daily 1:30 tours and 98 specially-arranged tours and programs. In FY2010 there were 5,506 visitors from the following 52 countries:

Afghanistan, Australia, Austria, Belgium, Brazil, Brunei, Burma, Cameroon, Canada, Cambodia, China, Colombia, Democratic Republic of Congo, Egypt, France, Germany, Ghana, Haiti, India, Indonesia, Israel, Italy, Japan, Kenya, Latvia, Laos, Lebanon, Malaysia, Mali, Mexico, Morocco, Nepal, Netherlands, Nigeria, Paraguay, Philippines, Poland, Republic of South Africa, Romania, Russia, Saudi Arabia, Sweden, Slovenia, Taiwan, Tanzania, Turkey, Uganda, United Kingdom, Ukraine, Uruguay, Venezuela, Vietnam

LIBRARY OPERATIONS

Sheldon Kotzin
Associate Director

The National Library of Medicine 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.

LO 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 (BSD), Public Services (PSD), Technical Services (TSD), and History of Medicine (HMD); three units: the Medical Subject Headings (MeSH) Section, the National Network of Libraries of Medicine (NN/LM) National Network Office (NNO), and the National Information Center on Health Services Research and Health Care Technology (NICHSR); and a small administrative staff. The activities of all these components receive essential support from a wide range of contractors.

Most LO activities are critically dependent on automated systems developed and maintained by NLM's Office of Computer and Communications Systems (OCCS), National Center for Biotechnology Information (NCBI), or Lister Hill National Center for Biomedical Communications (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 FY2010, LO published a Strategic Plan for 2010-2015. The Strategic Plan process involved numerous LO staff and many hours of research, discussion, and deliberation on the part of three Working Groups: Bibliographic Control; Collection, Preservation and Access to Information; and Workforce for the Future. The efforts and reports produced by these three groups form the basis of the identified priorities, goals and objectives incorporated in the Strategic Plan. These were subsequently organized into three strategic directions which will receive focus over the coming five years: Transforming Access to the Collection; Redesigning Systems and Workflows; and Developing a 21st Century Workforce. Additional resources will be sought to support them, as needed. Groups have been formed for each of the 10 strategic priorities within these three overarching areas and staff appointed to lead them.

In FY2010, LO continued to review and revise policies, procedures, services, and organizational lines to address shifting workloads, to use electronic information to enhance basic operations and services, and to work with other NLM program areas to ensure permanent access to electronic information through the development of an NLM digital repository. The digital repository, Digital Collections, was released providing rich search, browse and cross-collection access to the first two collections containing 518 cholera monographs and 11 historic public health films. Work on the repository involved a number of LO and OCCS staff. NLM/LO received a grant from the Sloan Foundation to be a partner in the Medical Heritage Library (MHL) project. Work also continued on the Indexing 2015 initiative, an NLM-wide research and development effort to improve indexing performance and productivity which is being led by LO. Both projects will be discussed in more detail later in this report. LO continued to develop its Continuity of Operations Plan to keep essential services operating in the event of a disaster.

After a thorough assessment of the MedlinePlus Go Local program, including extensive consultation with and feedback from Go Local sites, NLM made the decision to cease supporting it in February 2010. The main reasons for this decision were continued low use and the high cost of maintenance to NLM and the libraries and organizations sponsoring sites. By the end of FY2010 only three NLM-hosted sites (AL, IN, MI) and two locally hosted sites (NC, SC) remained. With the release of the MedlinePlus

redesign in July, MedlinePlus no longer links to Go Local sites.

Although many LO efforts are devoted to dealing with electronic information and supporting NLM's high priority outreach initiatives, LO must also devote substantial resources and attention to the care and handling of physical library materials and the space and environment for staff, patrons, and physical and electronic collections. Some areas of the collection are out of space and the remaining space will be completely filled by the end of 2010. While NLM's plans for a new building include increased space for the growing collection, funds have not yet been appropriated for the building. LO, with input from the Office of Administrative Management Analysis and Services (OAMAS), developed a plan several years ago for expanding existing space. The plan involves strengthening the B2 floor in the collection area as well as the ceiling of B3 below, installing new sprinklers, lighting, and HVAC, and installing compact shelving in phases over several years. Implementation of the plan will accommodate collection growth until approximately 2030. In FY2010 renovation of the southwest corner of B3 was completed and 12,000 linear feet of compact shelving was installed. The Accession Number monographs were moved from the southwest to the northwest corner of B2 and conventional shelving was removed from the southwest corner in preparation for floor strengthening and installation of new lighting, sprinklers and compact shelving rails.

In FY2010, LO's Administrative Office continued to provide guidance and assist managers, supervisors, and staff with a wide range of cross-functional administrative requirements, including acquisition, budget, human resources and other key areas. During the past year, the Administrative Office began phasing in several fundamental changes to day-to-day business practices including regular monthly meetings with each Division, management reports, and assigned roles for the administrative staff who service LO. Library Operations continued to encourage staff to take advantage of flexiplace work arrangements as appropriate. More than 170 LO employees work at home at least one day per week.

Collection Development and Management

NLM's comprehensive collection of biomedical literature is the foundation for many of the Library's services. LO ensures that this collection meets the needs of current and future users by updating NLM's literature selection policy; acquiring and processing relevant literature in all languages and formats; organizing and maintaining the collection to facilitate current use; and preserving the content for subsequent generations. At the end of FY2010, the NLM collection contained 2,663,158 volumes and 15,278,210 other physical items, including manuscripts, microforms, pictures, audiovisuals, and electronic media.

Selection

Selectors worked on a variety of projects to enhance the breadth and depth of the NLM collection. Work continued on identifying and acquiring grey literature on disaster management and public health. In addition a significant number of new grey literature titles in environmental health and global health issues were also identified. Audiovisuals of historical importance from the World Health Organization on tropical diseases and from the Centers for Disease Control and Prevention on black lung, influenza, and radiation emergency preparedness were selected for the collection. NLM received a donation of Japanese books from the University of California, Berkeley's East Asian Library which HMD and TSD continue to review for addition to the collection. Gifts were received from the State Library of Iowa, Brooke Army Medical Center Medical Library, Dwight D. Eisenhower Army Medical Library, NASA Kennedy Space Center Library, Mid-Pacific Region U.S. Bureau of Reclamation, Smithsonian Institution Libraries, and E. H. Munro Medical Library, St. Mary's Hospital, Grand Junction, CO resulting in the addition of 104 volumes to the NLM collection.

Acquisitions

TSD received and processed 140,261 contemporary physical items (books, serial issues, audiovisuals, and electronic media) which is slightly higher than last year's total. Electronic publishing has not yet had a significant impact on the number of physical items that NLM acquires. The number of serial issues processed declined by 7.7% for several reasons. The number of duplicate issues processed decreased 20% to 23,418 issues, mainly due to efforts to contact publishers and ask them not to send duplicate issues of journals for indexing. The number of journals that have stopped producing a print version in addition to an online version increased from 342 in FY2009 to 432 in FY2010. Finally, efforts continued to reduce serial expenditures by cancelling some titles that were duplicates, office copies, or out of scope for the collection. Net totals of 35,300 volumes and 3,532,323 other items, including nonprint media, manuscripts, and pictures acquired by the HMD, were added to the NLM collection.

LO uses subscription agents and book vendors to acquire current literature published around the world. In March 2010, new five-year Blanket Purchase Agreements for Domestic and Foreign Monographs and Continuations were awarded to five different companies.

HMD acquired a wide variety of important printed books, manuscripts and modern archives, images, and historical films during FY2010. Among them was Dalechamps, Jacques. *Chirurgie Francoise* (Lyon, 1569). One of the great rarities of the surgical literature, this manual, along with Ambroise Pare's *Dix Livres de Chirurgie* (1546) is considered one of the most important

French surgical books of the 16th century. The 1569 edition is so rare that many scholars have cited the 1570 edition as the first edition. The Library also acquired Petrus de Avano. *De venenis.* (Leipzig: Jacob Thanner, 14 July 1498). This book is a rare later printing of the first printed book on toxicology. A rare incunable edition of Razi, Abu Bakr Muhammad ibn Zakariya', *Opera,* (Venice, 1500) was acquired. It includes the "Liber ad Almansoris," a short general textbook on medicine and is one of the most widely read and influential medieval medical works.

Recently acquired archives and modern manuscript collections included over 60 linear feet of materials from Dr. Harold Varmus, former Director, NIH and Memorial Sloan-Kettering Cancer Center, relating to his tenure at both institutions; the papers of Dr. William DeVries, the surgeon who implanted the first total mechanical heart in Barney Clark; six boxes of medical ephemera donated by William Helfand; films on epidemiology from Dr. Martine Joan Work; and films of Ivan Pavlov's lab received from Richard Foa.

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 began a multi-year inventory of the serials collection in FY2006. By the end of FY2010, the contractor, CBase Solutions Inc., had inventoried 82,025 serial titles. Among the completed titles are all journals indexed in *Index Medicus* and *Index Catalogue*, all currently received English-language titles, and the top 500 non-IM/MEDLINE titles requested on interlibrary loan via the DOCLINE system. The inventory, to be completed in early FY2011, will make NLM's collection more accessible by providing accurate holdings information at the issue level to staff and onsite users as well as to libraries and individuals worldwide. Offers were received from 196 libraries through the Journal Donation Program to fill gaps in the NLM collection. A total of 3,228 gift volumes or issues were added to the collection, a 72% increase over the previous year.

In FY2010, LO bound 19,629 volumes, repaired 2,727 items in the onsite repair and conservation laboratory, made 441 preservation copies of films and audiovisuals, and conserved 4,125 items. A total of 551,660 items were shelved. This figure is a 7% decrease from the FY2009 total of 590,419 items shelved and reflects the decline in interlibrary loan and Main Reading Room requests.

Permanent Access to Electronic Information

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

PubMed Central (PMC), a digital archive of medical and life sciences journal literature developed by NCBI, is NLM's vehicle for ensuring permanent access to electronic journals and digitized 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's Public Services Division (PSD) continued to work closely with NCBI to scan and add digitized back files of journals depositing newly published articles in the archive. PSD prepares back issues for scanning, ships them to the scanning contractor, and manages the human review portion of the quality control of scanned images, accompanying OCR data, and XML-tagged citations for articles that pre-date current MEDLINE/PubMed coverage. Because the bindings are cut to make scanning more efficient, NLM does not use volumes from its archival collection in this effort. Instead, NLM solicits copies from publishers and other libraries. In FY2010, three new titles were added: *Epidemiology and Infection*, *The Journal of Hygiene,* and *The Analysis of Verbal Behavior.* By the end of FY2010, 1,207,474 scanned articles and 2,054,345 articles in total were in the database.

Continuing work begun in FY2006 on the development of an NLM digital repository, Digital Collections was launched in September 2010, providing rich search, browse and access within and across two collections—518 cholera monographs and 11 historical public health films.

The NLM is one of five institutions (NLM, Yale, Harvard, Columbia and New York Public Library) participating in a project to create the Medical Heritage Library (MHL) by digitizing nearly 30,000 volumes of historical medical materials. The project is funded by the Alfred P. Sloan Foundation for a two-year period (January 2010-January 2012). NLM plans to create high quality, long-term digital surrogates of approximately 5700 rare and fragile items from the Americana collection published between 1720 and 1865 that comprise an estimated 1.7 million page images. Under the oversight of the Deputy Associate Director of LO, a cross-divisional team of staff from HMD, PSD, TSD and OCCS are working on the project which is Library Operations' first effort to develop in-house capacity to digitize its collections for long term access and preservation. In FY2010, a digitization lab was created, Kirtas scanning equipment was installed, a

workflow was developed, staff was trained and by the end of September digitization of 294 books was completed.

Vocabulary Development and Standards

LO produces and maintains the Medical Subject Headings (MeSH), a subject thesaurus used by NLM and many other institutions to describe the subject content of biomedical literature and other types of information; develops, supports, or licenses for U.S. use vocabularies designed for use in electronic health records and clinical decision support systems; and works with OCCS and LHNCBC 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 U.S. health data.

In FY2010 work was completed on the 2011 edition of MeSH which contains 26,142 descriptors and more than 197,479 supplementary records for chemicals and other substances. The MeSH Section added 573 new descriptors, replaced 54 descriptors with more up-to-date terminology and deleted 19 descriptors. Of special note are developments related to algae and sex disorders. Taxonomically algae are polyphyletic and therefore the descriptor Algae no longer easily fits into the MeSH trees. So for 2011 MeSH, the descriptor Algae was deleted and its children distributed among the appropriate eukaryotic trees. Based on the updated classification and new nomenclature recommendations put forth by the 2006 International Intersex Consensus Conference 1, the Disorders of Sex Development (previously Sex Differentiation Disorders) trees and descriptors were revised and updated. Work continued on merging the list of rare disease terms maintained by the Office of Rare Diseases Research (ORDR) into the MeSH vocabulary. The rare disease terms that matched existing MeSH descriptors were merged with those descriptors. The remaining terms were introduced as Supplementary Concept Records (Class 3) in MeSH 2011. Because rare diseases are defined as having a prevalence of fewer than 200,000 affected individuals in the United States they traditionally receive less attention and are sometimes called orphan diseases. Having these terms available for indexing purposes will enable a more precise retrieval of the rare disease articles and contribute to their identification.

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 have intensified, particularly in the areas of quality and performance measurement, lab services, and newborn screening.

The Library works closely with the office of the National Coordinator for Health Information Technology. Staff members of LO's National Information Center on Health Services Research and Health Care Technology (NICHSR) routinely contribute to various efforts related to standards development and interoperability. During FY2010, one of the new Federal oversight bodies, the Health IT Standards Committee, established a Vocabulary Task Force (VTF). Co-Chaired by NLM's Deputy Director, the VTF members include representation from MeSH and LHNCBC; NICHSR staff provided support to this effort in 2010. In addition, NICHSR staff members routinely monitor the Federal Health Architecture (FHA) initiative; participate in the Public Health Data Standards Consortium; serve on the HHS Data Council; serve as staff to the National Committee on Vital and Health Statistics (NCVHS) Standards Subcommittee; and with AHRQ to support efforts to promote patient safety and adverse event reporting. In FY2010, NICHSR participated in the National Quality Forum's Code Maintenance Project, established to advise the NQF on how to maintain its endorsed quality and performance measures during the upcoming required migration of coding from ICD-9-CM to ICD-10-CM/PCS and SNOMED CT. In FY2010, NICHSR continued to represent NLM and participate on the AHRQ Expert Panel for the National Guidelines and the National Quality Measures Clearinghouses. All of these efforts are designed to promote standardized data collection and reporting on quality of care and improvement activities.

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 2010 took place in April and November.

Clinical Vocabularies

The MeSH Section and its contractors also produce RxNorm, a clinical drug vocabulary that provides standardized names for use in prescribing. It is released through the UMLS. RxNorm was designated as a U.S. government-wide target clinical vocabulary standard by the Secretary of the Department of Health and Human Services as one of a suite of standards for use in U.S. federal government systems for the electronic exchange of clinical health information. It represents the information that is typically known when a drug is prescribed, rather than the specific product and packaging details that are available at the time a medication is purchased or administered, and provides a mechanism for connecting information from different commercial drug information services. In FY2010, RxNorm editors prepared and released monthly updates to the clinical drug vocabulary. NLM and the Department of Veterans Affairs, Veterans Health Administration signed an agreement making it possible for the VHA National Drug File – Reference Terminology (NDF-RT) to be included as a source vocabulary in RxNorm. 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.

Through NICHSR, NLM uses a contract mechanism to support the continued development and free distribution of LOINC (Logical Observation Identifiers Names and Codes) by the Regenstrief Institute. LOINC is a clinical terminology important for laboratory test orders and results. In 2003, LOINC was designated by the Secretary of the Department of Health and Human Services as one of a suite of standards for use in U.S. Federal government systems for the electronic exchange of clinical health information. In 2005, the Secretary proposed adoption of LOINC as a HIPAA standard for some segments of Claims Attachment transactions. In 2007, LOINC was identified as a standard in several HITSP Implementation Specifications. In 2010, LOINC was identified as one of several required standards for use in EHRs to achieve meaningful use, as defined by the ARRA-HITECH provisions. In FY2010, NLM continued its support for LOINC.

In FY2010, NLM continued to support and pay the annual fees for the U.S.-wide license for the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT). SNOMED CT is a comprehensive clinical terminology. In 2004, SNOMED CT was designated by the Secretary of HHS as one of the suite of standards for use in U.S. Federal Government systems for the electronic exchange of clinical health information. In 2007, SNOMED CT was identified as a standard in several HITSP Implementation Specifications. In 2010, SNOMED CT was identified as one of several required standards for use in EHRs to achieve meaningful use, as defined by the ARRA-HITECH provisions.

SNOMED CT is owned by the International Health Terminology Standards Development Organisation (IHTSDO). NLM, on behalf of HHS, is the U.S. member of the IHTSDO responsible for distribution of SNOMED CT within the U.S. and for representing U.S. interests in the continued development of SNOMED CT. In October 2010, NLM hosted the semi-annual conference of the Members and Affiliates of the IHTSDO, an international gathering that drew participants from Europe, Asia and South America. In 2010, NLM hired a new SNOMED CT clinical expert (in MeSH). This new hire enabled NLM, with participation from MeSH, MMS, NICHSR, and OCCS, to begin work on the U.S. SNOMED CT Request Submission System and development of the U.S. Extension to SNOMED CT, both of which will be available sometime in FY2011. NLM continues working with the IHTSDO to facilitate negotiations for the alignment and harmonization between SNOMED CT and other key health terminologies, including LOINC and RxNorm.

NLM continues to be involved in activities related to the development of mappings between important standard vocabularies (SNOMED CT and LOINC) and administrative code sets (such as ICD-9-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. Achieving this goal will require that key clinical data elements are captured or recorded in detailed, standardized form (using standard vocabularies, codes, and formats) as close to their original sources (patients, health care providers, laboratories, diagnostic devices, etc.) as possible. If these standardized clinical data can also be used to generate HIPAA-compliant billing transactions automatically, this will provide another incentive for adoption of clinical data standards. For automated generation of bills from clinical data to become a reality, robust mappings from standard clinical terminologies to the HIPAA code sets must be created. HHS has given NLM the responsibility for funding, coordinating, and/or performing official mappings between standard clinical terminologies and HIPAA code sets. In FY2010, NLM (including NICHSR, MeSH and LHNCBC staff) continued to participate in several mapping projects, both nationally and internationally, to achieve these goals.

Bibliographic Control

LO produces authoritative indexing and cataloging records for journal articles, books, serial titles, films, pictures, manuscripts, and electronic resources, using MeSH to describe their subject content. LO also maintains the NLM Classification, a scheme for arranging physical library collections by subject that is used by health sciences libraries worldwide. NLM's authoritative bibliographic

data improve access to the biomedical literature in the Library's own collection, in thousands of other libraries, and in many electronic full-text repositories.

Cataloging

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

RDA (Resource Description and Access) is the new set of cataloging rules intended to replace the Anglo-American Cataloguing Rules, 2nd edition (AACR2). Among RDA's aims is to better embrace a growing and international community of metadata users and to permit integration with emerging database technologies as well as the Semantic Web. Concerns were raised in 2008 about how well RDA met its own self-stated goals, the costs of adopting new rules, and the lack of clear explanations of the new rules' benefits. Therefore, the U.S. community, led by the senior management at the NLM, the National Agricultural Library (NAL), and the Library of Congress (LC), agreed to coordinate a U.S. test of the rules to assist them in making a joint decision about whether or not the rules should be adopted. The U.S. RDA Test Coordinating Committee was appointed in June 2008, with members from NLM, NAL and the LC. In addition to the three national libraries, 23 other test partners of various types and sizes of libraries throughout the U.S. planned to test the new rules. The goal of the test was to assure the operational, technical, and economic feasibility of RDA.

During FY2010 the Coordinating Committee members from LC, NAL and NLM finalized the U.S. RDA test procedures and survey instruments and began working on defining evaluative measures. The test period officially began on July 1, with the first three months allotted to become familiar with the online tool and create practice records. The participating catalogers and supervisory staff attended biweekly meetings to discuss cataloging and policy decisions regarding RDA. In FY2011 the three national libraries will make a joint decision on whether or not to implement RDA, based on the results of the test.

In FY2010, the Cataloging Section cataloged 18,820 books, serial titles, nonprint items, and cataloging-in-publication galleys. However, this represents a 9% decrease in items cataloged from the previous year's production figures. Several factors account for the decreased cataloging production figures for FY2010. The Section had two vacant cataloger positions until late in the fiscal year. In addition, three experienced catalogers were involved in the RDA testing process which had a negative impact on their production levels. Finally, several work days were lost in the aftermath of the blizzards that struck the area in February.

Progress was made in providing cataloging records for NLM's historical and special collections. HMD cataloged or upgraded catalog records for 4,328 early monographs, 2018 linear feet of manuscripts, 14,751 pictures, and 290 audiovisuals.

The Cataloging Section released the 2010 online edition of the massive *NLM Classification*. The WX (Hospitals and Other Health Facilities) and WY (Nursing) schedules were reviewed. Cataloging has been using the Medical Text Indexer (MTI) developed by the Lister Hill Center to suggest subject headings for cataloging records. In FY2010, the Cataloging Section began a project to enhance mapping of Library of Congress Subject Headings (LCSH) to MeSH using a list of LCSH terms found on records imported into Voyager that are not already mapped in MTI. Catalogers also began scanning the weekly lists of new LCSH, mapping the relevant headings to MeSH and forwarding them to LHC for inclusion in MTI.

Indexing

LO indexes 5,484 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 U.S. and international institutions indexed 699,420 articles in FY2010, bringing the total number of MEDLINE citations to over 18 million. Indexed citations were updated to reflect 243 retracted articles, 8,433 published corrections, and 35,688 comments found in subsequently published notices or articles.

In FY2010, indexers created 80,042 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI Entrez Gene database. This was a 3.5% decrease 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. Phases I and II of the Medical Text Indexer (MTI) experiments, part of the Indexing 2015 initiative, to test use of MTI to assist in the indexing process were completed. These two phases focused on microbiology journals. The results were so promising that a Phase III test was planned for journals in other subject areas and BSD staff made the recommendation to implement MTI as a "front-line indexer" to be revised by a human reviser for a select number of journals in the first half of FY2011. The Index Section continues to increase the number of articles indexed from the online version of journals. At the end of FY2010, 3,758 journals were indexed from an online version, including online-only journals and those with a print version. Freeing the print version for immediate use onsite and for fulfilling interlibrary loan requests, indexing from an online version has also enabled the elimination of

duplicate print issues, saving the cost of processing and shelving those issues.

Indexers perform their work after the initial data entry of citations and abstracts has been accomplished. Over the past 10 years, great strides have been made in improving the efficiency of data entry. By the end of FY2010, 90% of all citation data entry consisted of XML-submitted data from publishers, a 2% increase over FY2009. 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 FY2010, LSTRC reviewed 528 journals and rated 132 of them highly enough for NLM to begin indexing them.

Information Products

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

Databases

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

Use of MEDLINE/PubMed, which now includes over 20 million citations, registered 1.58 billion searches in FY2010 which represents a 23% increase from the previous year. Additional progress was made in adding or enhancing citations to older articles; 76% percent of OLDMEDLINE subset citations have original key words mapped to current MeSH. BSD staff also assisted NCBI with the design, development, and testing of many enhancements to PubMed, notably: redesign of the interface including a new homepage; streamlining and redesigning the PubMed Advanced Search page to include a search builder function, moving Limits to its own page as well as many other changes and improvements.

Use of MedlinePlus increased to 153 million unique visitors over 123 million unique visitors in FY2009; MedlinePlus en español saw a similar increase, with 63.8 million unique visitors in FY2010, an increase from 46.9 million unique visitors in FY2009. MedlinePlus and MedlinePlus en español continue to receive high ratings from customers in the American Customer Satisfaction Index (ASCI), ranking among the top government news/information sites.

PSD and OCCS continued to expand and improve the content and features of the English and Spanish versions of MedlinePlus. A redesign of the site with updated graphics, layout and architecture was released. At the end of FY2010, the site featured 859 health topics in English and 826 in Spanish. A collection of patient handouts, which provide short, printable health information, was added in response to user requests. The collection appears as a subcategory on the health topic pages and was released with more than 1,000 links in English and nearly 500 in Spanish. Mobile MedlinePlus, which is a mobile Web site making it generic for use on any mobile device, was released in January. A pilot service that connects electronic health, medical or patient records with context-specific health topics and drug monographs on MedlinePlus called MedlinePlus Connect, was launched in collaboration with the Institute for Family Health, Cleveland Clinic and Columbia University. GovDelivery, an email update service that allows users to customize their subscriptions to content was also launched. Subscribers to the MedlinePlus English and Spanish content updates via GovDelivery now number more than 87,000 with subscriptions to content totaling 5.3 million.

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 FY2010. NICHSR worked with NCBI to add Centers for Disease Control and Prevention publication *Health US 2009* to the Entrez Bookshelf; NCBI also continued to expand the set of clinical guidelines-related materials on Bookshelf drawing from such sources as the National Institute for Health and Clinical Excellence (NICE) and the Institute of Medicine. In FY2010, NICHSR participated in discussions with NCBI on ways in which NLM resources could support comparative effectiveness research. NICHSR also continued to work with NCBI to improve the HSTAT (Health Services and Technology Assessment Text) on the Bookshelf , including the addition of 50 new documents produced by the Agency for Healthcare Research and Quality, primarily evidence reports, technology assessments, and evidence syntheses.
In FY2010, 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), both visits and unique visitors increased more than 40% from the end of FY2009 to the end of FY2010. Page views increased more than 30% over the same time period.

NICHSR continued to contribute to the field of health services research through its support of NLM databases containing information about newly-funded research and accessible data collection tools and resources. With the assistance of AcademyHealth and the Sheps Center at the University of North Carolina, Chapel Hill, the content of HSRProj (Health Services Research Projects in Progress) continued to expand, incorporating work funded by additional foundations, states, and other organizations. In 2010, the database grew to more than 8000 active records representing more than 110 funders and was increasingly recognized as a valuable database in its own right and used to monitor projects related to comparative effectiveness, and other topics.

The Health Services Research Resources (HSRR) database also continued to expand to cover additional datasets, surveys, other research instruments, and software packages used with datasets. In addition to working with the librarian, research, and policy communities to improve the usability and value of its HSR web offerings, NICHSR, working in conjunction with AcademyHealth and the Medical Library Association, sponsored and conducted a two-part webinar series on *Making the Most of Research Resources for HSR*, designed for a mixed audience of library and information professionals as well as health policy and health services researchers interested in learning more about research resources for topics in health services, delivery, and outcomes.

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.

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 FY2010, there were 621 licensees of MEDLINE data, a 17% increase over the previous year. The majority use the data for research and data-mining.

At the end of FY2010, there were 6,116 UMLS licensees, an increase of 26% over the previous year. NLM services and support to licensees of UMLS were enhanced by two webcasts and expanded UMLS Source Release Documentation to include information on how original source information is transformed and represented in the UMLS Metathesaurus.

Web and Print Publications

NLM's 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. BSD's MEDLARS Management Section edits the *NLM Technical Bulletin*, which provides timely, detailed information about changes and additions to NLM's databases and related policies, primarily for librarians and other information professionals. Published since 1969, the *Technical Bulletin* also serves as the historical record of the evolution of NLM's online systems and databases.

NLM added one new segment to the *Profiles in Science* Web site during FY2010, bringing the total to 34. The new site focuses on Daniel Nathans in collaboration with the Allen M. Chesney Archives of the Johns Hopkins Medical Institutions, which holds the Nathans papers. Nathans was an American molecular biologist who worked with restriction enzymes. Launched in September 1998, *Profiles in Science* promotes the use of the Internet for research and teaching in the history of biomedical science by making widely available archival collections of leaders in biomedical research and public health. Published and unpublished materials appear on the site, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audiotapes, and video clips. Altogether the program has digitized more than 175,000 pages of material.

In FY2010, 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. Three issues of The *NIH MedlinePlus Magazine* were published in print and online in FY2010.

Direct User Services

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

Document Delivery

LO retrieves documents requested by onsite patrons from NLM's closed stacks and also provides interlibrary loan as a backup to document delivery services available from other libraries and information suppliers. In FY2010,

PSD's Collection Access Section processed 401,699 requests for contemporary documents, a 9% decrease from FY2009. HMD handled 7,676 requests for rare books, manuscripts, pictures, and historical audiovisuals.

The number of onsite users registering to use the collection continued to decline by another 11% from last year and use of NLM's collection by users in the Main Reading Room was also down 12% from the previous year's total to 145,240. Users of the HMD Reading Room requested 6,901 items from the historical and special collections, an increase of 1% from last year. Paid printing at Main Reading Room workstations decreased 26%, from 213,215 pages in FY2009 to 156,780 pages in FY2010, a trend that reflects the increased use of flash drives to download electronic journal content.

The Collection Access Section (CAS) received 256,459 interlibrary loan requests, a 4% decline from FY2009 with a fill rate of 83%. The number of requests processed in 12 hours increased slightly to 97%, with 98% processed within one day of receipt. NLM continues to deliver 96% of interlibrary loan requests electronically. CAS reorganized the Interlibrary Loan Unit by combining the Serials and Monographs/AV Groups into the single ILL Processing Group.

A total of 2,904 libraries use DOCLINE, NLM's interlibrary loan request and routing system. DOCLINE users entered 1,765,108 requests in FY2010, a 5% decline from last year; 93% of requests were filled. Individuals submitted 329,682 document requests to DOCLINE libraries via the Loansome Doc feature in MEDLINE/PubMed and the NLM Gateway, an 11% decline from the previous year. Document request traffic continues to decline in all Regions of the NN/LM due to expanded availability of electronic full text journals.

NCBI and the staff at the Regional Medical Libraries continued to support and promote the use of PubMed's LinkOut for Libraries and Outside Tool, the open-URL services that allow libraries to link directly from PubMed to a wide range of resources beyond the Entrez system. Using these tools, libraries can create custom displays of their electronic and print holdings for their primary clientele. The number of libraries participating in LinkOut increased by 8.5% in FY2010 to 2,415; there are 750 libraries participating in the Outside Tool option, an increase of 16% over last year. A LinkOut for Libraries Training and Education Resources Web site was launched which includes links to all LinkOut related Quick Tours, LinkOut User Meeting slides from MLA 2010, and links to other resources of interest to LinkOut Libraries.

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

Reference and Customer Services

LO provides reference and research assistance to onsite and remote users as a backup to services available from other health sciences and public libraries. LO also has primary responsibility for responding to inquiries about NLM's products and services and how to use them effectively. LO's Reference and Web Services Section responds to initial inquiries and also handles the majority of questions requiring second-level attention. staff from throughout LO and NLM assist with second-level service when their special expertise is required. A total of 98,616 inquiries were received in FY2010, up 9% from FY2009. The number of onsite inquires declined slightly to 7,076. The number of remote inquiries increased 9% to 91,540 with the overwhelming majority arriving via e-mail.

Outreach

LO manages or contributes to many programs designed to increase awareness and use of NLM's collections, programs, and services by librarians and other health information professionals, historians, researchers, educators, health professionals, and the general public. LO coordinates the National Network of Libraries of Medicine which attempts to equalize access to health information services and information technology throughout the United States; serves as secretariat for the Partners in Information Access for the Public Health Workforce; participates in NLM-wide efforts to develop and evaluate outreach programs for underserved minorities and the general public; produces major exhibitions and other special programs in the History of Medicine Division; 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 U.S. health professionals, researchers, and the general public irrespective of their geographic location. With more than 5,964 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 FY2010, a major initiative was the continued development of a national NN/LM Emergency Preparedness Plan. Activities under this initiative included the development of a toolkit for use by network libraries in preparing for and responding to emergencies. The toolkit was evaluated for usability and recommendations were incorporated into a revised organizational structure and improved design.

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 FY2010, the NN/LM initiated 220 outreach projects which target rural and inner city communities and special populations in 39 states, the District of Columbia and the U.S. Virgin Islands.

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, D.C. area, and some distant meetings focused on health services research, public health, and history of medicine. In FY2010, NLM and NN/LM services were exhibited at 41 national and 374 regional, state, and local conferences across the U.S. These exhibits highlight all NLM services relevant to attendees.

Partners in Information Access for the Public Health Workforce

The NN/LM is a key member of the Partners in Information Access for the Public Health Workforce, a 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. In FY2010, the National Association of Local Boards of Health (NALBOH) joined the Partners.

The Partners Web site (http://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 FY2010, the Web site was expanded with more than 640 new links. Two new topic pages on Veterinary Public Health and How to Access Journal Articles were also added to the Web site. Usage of the site continued to increase in FY2010; comparing fourth quarter 2010 to the same period in FY2009, visits increased 26%, unique visitors more than 20% and page views more than doubled.

One of the most popular resources on the PHPartners site is the Healthy People 2010 Information Access Project (HP2010 IAP), which provides structured PubMed searches and links to MedlinePlus for a selected set of objectives included in HP 2010. In FY2010, NICHSR 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. NICHSR contracted with 20 librarians to develop Healthy People/PubMed Structured Evidence Queries (HP-SEQs, called “HP-seeks”) and anticipates these searches will be incorporated into HP by mid-2011.

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 Director 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 and Nigeria were joined by a librarian from Morocco as a Fellow in the FY2010-2011 Associate Program. 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 Director 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. African journals that have been included in this project over the past several years are: African Health Sciences, Ghana Medical Journal, Malawi Medical Journal, Mali Medical, Ethiopian Journal of Health Sciences, and Medical Journal of Zambia.

LO also continued to provide support to NLM’s joint fellowship program with the Association of Health Care Journalists, now in its second year. In FY2010, NICHSR again presented to the journalists and AHJC staff on health services research and public health resources

available through NLM and its partner agencies, and the Administration's OpenGov Initiative.

Historical Exhibitions and Programs

HMD directs the development and installation of major historical exhibitions in the NLM rotunda, with assistance from LHNCBC and the Office of the Director. Designed to appeal to the interested public as well as the specialist, these exhibitions highlight the Library's historical resources and are an important part of NLM's outreach program. The current exhibition on display, *Against the Odds: Making a Difference in Global Health,* examines the revolution in global health that is taking place in towns and cities around the world. It presents a look at the public health problems posed by Hurricane Katrina. It showcases the barefoot doctors program, which trained over one million young people to treat the common ailments of residents of rural China in the 1960s and 1970s. The exhibition also profiles a campaign for oral rehydration in Bangladesh that was so successful that it has been adopted in Afghanistan as well. In another example of nation-to-nation collaboration, *Against the Odds* shows how the Pholela Health Center in South Africa inspired the community health center movement in the U.S.

Targeting local communities, announcements about *Against the Odds* were posted on the Bethesda Urban Partnership Calendar. In addition, a teacher development day for five Prince George's County educators was held and Exhibition Program staff attended a vendor fair for Howard County Public Schools to share information about the Exhibition Program and NLM's resources with Howard County science educators. Information about the NLM exhibitions was also shared with Frederick County Public School teachers and with educators in Loudon and Fairfax County in Virginia.

Previous NLM exhibitions live on through heavily used Web sites, printed catalogs, DVDs, and touring traveling versions. The travelling version of *Changing the Face of Medicine: Celebrating America's Women Physicians*, funded by the NIH Office of Research on Women's Health and NLM, continued touring libraries in the U.S. through a collaborative arrangement with the American Library Association. Additional travelling exhibitions that are touring include: *Against the Odds: Making a Difference in Global Health; Everyday Miracles: Medical Imagery in Ex-Votos; Frankenstein: Penetrating the Secrets of Nature; Harry Potter's World: Renaissance, Science, Magic and Medicine; Opening Doors: Contemporary African American Academic Surgeons; Literature of Prescription: Charlotte Perkins Gilman and the Yellow Wall-Paper; and Rewriting the Book of Nature: Charles Darwin and Evolutionary Theory.*

Besides the major exhibitions mounted in the rotunda, HMD installed several special displays. *The Henkel Physicians: A Family's Life in Letters* featured selections from the Henkel manuscript collection and was accompanied by the online launch of the *Physicians in the Shenandoah Valley* exhibition. *Nothing to Work With But Cleanliness: African American "Grannies," Midwives & Health Reform* focused on African American nurse midwives. The first edition of Charles Darwin's *On the Origin of Species* (1859) was placed on display in the exhibition, *Rewriting the Book of Nature: Charles Darwin and Evolutionary Theory* in connection with the symposium, Finished Proofs on October 1, the anniversary of Darwin's announcement that he had finished proof reading his book. A banner exhibition in the Lister Hill Center lobby commemorating African American History Month, *Within These Walls: the Contraband Hospital and the African Americans Who Served There* highlighted the achievements of African American medical workers at a hospital during the Civil War.

In a novel twist on NLM's popular online system, *Turning The Pages*, http://archive.nlm.nih.gov/proj/ttp/books.htm), which permits users to turn the pages of a rare book on their computer screen, journeyed to pre-book times to allow them to "unroll the scroll" or, more specifically, the Edwin Smith Papyrus, the world's oldest known surgical document. The Smith Papyrus was written in Egyptian hieratic script around the 17th century BCE (Before the Common Era) but probably based on material from a thousand years earlier. This collaborative online representation features an important new translation by James P. Allen, formerly of the Metropolitan Museum of Art, and high-resolution scans lent by the scroll's owner, the New York Academy of Medicine. A new project was also released on the Turning the Pages kiosks in the Library's Visitor Center and the History of Medicine Reading Room, featuring Hanaoka Seishu's "Surgical Casebook," a magnificently illustrated manuscript depicting the likenesses of the men and women who came to him for treatment in early 19th century Japan. Hanaoka was a pioneering Japanese surgeon who was the first to use general anesthesia to remove tumors from cancer patients. The images are colorful, often charming, and depict quite graphically the medical and surgical problem to be treated.

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

Training and Recruitment of Health Sciences Librarians

LO staff 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 and Clearinghouse (NTCC) at the New York Academy of Medicine; 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, to promote multi-institution evaluation of library services, and to encourage specialist roles for health sciences librarians.

In FY2010, the MEDLARS Management Section (MMS) and the NTCC trained 1,026 students in 70 classes covering PubMed, the NLM Gateway/ClinicalTrials.gov, TOXNET, and the UMLS. This represents a 6% decrease from the number trained in FY2009.

The UMLS for Librarians course and the UMLS Tutorial continue to represent some of the NLM training courses useful in preparing librarians for new and expanded roles. Although NCBI made the decision to discontinue offering the "Introduction to Molecular Biology Information Resources" course, it was recorded for ongoing public use. NICHSR continues to add to its suite of courses on health services research, public health, and health policy.

In addition to the courses mentioned previously concerning use of NLM resources, in FY2010, NICHSR developed and conducted a four-part webinar series on health indicators targeted to the medical librarian community (for whom CE credits were available). Building on the class taught in 2009 at the MLA annual meeting, the course included an overview of the design and use of health indicators, an in-depth review of county-level indicator projects including the Community Health Status Indicators project, practical approaches for librarians in using health indicator data to engage with their communities, and an exploration of several local health indicator efforts in action. The course was well attended at the time and continued to attract viewers later in the year.

The NLM Associate Fellowship program had eight participants in FY2010: three second year fellows at sites across the country (the University of Minnesota, University of Texas at San Antonio, and the University of Utah) and five first year fellows, four of whom completed their year at NLM in August 2010. Second year placements were arranged for three of these fellows at the National Institutes of Health Library, Bethesda; University of Maryland, and Yale University and one fellow took a position at the health sciences library of Inova Fairfax Hospital. Six new fellows began the 2010/11 Associate fellowship year at NLM in September, including an international fellow from the National Institute of Hygiene, Rabat, Morocco. Efforts to recruit fellows from underrepresented groups have been successful in attracting diverse groups of fellows to the program, including Hispanic representation in the 2007/08 and 2008/09 cohorts and African American representation in 2009/10 cohort.

NLM works with several organizations on librarian recruitment and leadership development initiatives. Individuals from minority groups continue to be underrepresented in the library profession and a high percentage of current library leaders will retire within the next five to 10 years. LO has provided support for minority student scholarships available through the American Library Association, the Medical Library Association, and the Association for Research Libraries (ARL). In FY2010, LO provided funds to support the ALA Spectrum Scholarship Program for 10 more years. LO also supports the NLM/AAHSL Leadership Fellows Program which provides leadership training, exposure to critical issues, mentorship, and site visits to the mentor's institution for an annual cohort of five emerging leaders for academic health sciences library directorships. AAHSL contracts with consultants for program management and curriculum content. Recruitment efforts have emphasized and been successful in attracting minority candidates. Recruitment efforts have emphasized and been successful in attracting minority candidates. In FY2010, NLM continued participating in an ARL-directed Institute of Museum and Library Services (IMLS) grant to recruit minority students enrolled in graduate library and information science programs to work in research libraries. Three fellows spent six weeks at NLM during the summer.

Table 1. Growth of Collections

Collection

	Previous Total (9/30/09)	*Added FY2010*	*New Total (9/30/10)*
Book Materials			
Monographs:			
Before 1500	595	1	596
1501 – 1600	6,025	21	6,046
1601 – 1700	10,314	24	10,338
1701 - 1800	24,837	54	24,891
1801 – 1870	41,841	93	41,934
Americana	2,341	15	2,356
1871 – Present	820,390	12,855	833,245
Theses (historical)	288,091	0	288,091
Pamphlets	172,021	0	172,021
Bound serial volumes	1,389,473	23,418	1,412,891
Volumes withdrawn	(128,070)	(1,181)	(129,251)
Total volumes	2,627,858	35,300	2,663,158
Non-book Materials			
Microforms:			
Reels of microfilm	149,021	246	149,267
Number of microfiche	456,781	3	456,784
Total microforms	605,802	249	606,051
Audiovisuals	84,831	1,335	86,166
Computer software	2,560	0	2,560
Pictures	69,285	11	69,296
Manuscripts	10,984,382	3,530,800	14,515,182
Non-book items added	11,746,860	3,532,395	15,279,255
Non-book items withdrawn	(973)	(72)	(1,045)
Total non-book items	11,745,887	3,532,323	15,278,210
Total book & non-book items	14,373,745	3,567,623	17,941,368

Table 2. Acquisition Statistics

Acquisitions

	FY2008	*FY2009*	*FY2010*
Serial titles received	20,901	20,096	20,465
Publications processed:			
Serial pieces	138,893	129,053	119,067
Other	21,739	23,224	21,194
Total	160,632	152,277	140,261
Obligations for:			
Publications	$8,778,048	$9,549,598	$11,222,292
(For rare books)	($299,764)	($294,219)	($299,822)

NOTE: The FY2009 number of obligations reported last year was incorrect. It has been corrected in this report.

Table 3. Cataloging Statistics

	FY2008	*FY2009*	*FY2010*
Completed Cataloging	21,507	20,615	18,820

Table 4. Bibliographic Services

Services	***FY2007***	***FY2008***	***FY2009***
Citations published in MEDLINE	671,904	712,675	699,420
Journals indexed for MEDLINE	5,319	5,394	5,484
Journals indexed for Index Medicus	4,660	4,759	4,866
Total items archived in PubMed Central	1,683,664	1,869,309	2,052,693

Table 5. Consumer Web Services

Services	***FY2008***	***FY2009***	***FY2010***
NLM Web Home Page			
Page Views	54,600,000	42,700,000	44,500,000
Unique Visitors	11,700,000	7,900,000	7,700,000
MedlinePlus			
Page Views	754,000,000	660,000,000	716,000,000
Unique Visitors	132,300,000	123,000,000	153,000,000
ClinicalTrials.gov			
Page Views	563,956,116	785,589,163	857,270,086
Unique Visitors	10,466,607	10,794,930	10,583,773
DailyMed			
Page Views	16,436,561	66,546,000	70,195,000
Unique Visitors	1,688,000	1,902,000	3,035,000
Genetics Home Reference			
Page Views	27,252,298	36,465,215	118,267,551
Approximate Unique IPs	2,620,080	3,005,955	3,641,936
Household Products Database			
Page Views	15,142,059	140,340,909 *	10,984,725
Unique Visitors	1,054,447	875,237	850,332
Tox Town			
Page Views	5,039,964	6,352,651	6,096,928
Unique Visitors	224,762	300,230	286,004

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

Table 6. Circulation Statistics

Activity	*FY2008*	*FY2009*	*FY2010*
Requests Received	488,769	442,364	401,699
Interlibrary Loan	283,591	271,455	256,459
Onsite	205,178	170,919	145,240
Requests Filled	405,475	375,766	338,145
Interlibrary Loan	234,020	229,577	213,330
Onsite	171,455	146,189	124,815

Table 7. Online Searches—PubMed and NLM Gateway

(Data is processed completely differently from years prior to 2008, and should not appear with previous data. They are not comparable.)

PubMed Statistics	*FY2008*	*FY2009*	*FY2010*
Page Views	4,100,051,973	3,792,907,905	4,104,309,671
Interactive Sessions*	229,571,548	349,191,616	342,095,783
Searches	775,504,557	1,281,180,957	1,578,714,477

GateWay Statistics	*FY2008*	*FY2009*	*FY2010*
Page Views	6,391,604	7,794,098	8,467,087
Unique Visits	709,259	539,919	364,252
Searches	1,128,932	1,578,208	1,203,127

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

Table 8. Reference and Customer Services

Activity	*FY2008*	*FY2009*	*FY2010*
Offsite requests	85,457	83,140	91,540
Onsite requests	8,142	7,200	7,076
Total	93,599	90,340	98,616

Table 9. Preservation Activities

Activity	*FY2008*	*FY2009*	*FY2010*
Volumes bound	18,583	18,866	19,629
Volumes repaired onsite	2,412	3,174	2,727
Audiovisuals preserved	544	268	441
Historical volumes conserved	82	185	4,125

Table 10. History of Medicine Activities

Activity	*FY2008*	*FY2009*	*FY2010*
Acquisitions:			
Books	583	592	915
Modern manuscripts	820 (in ft)	2,776 (in ft)	680 (in ft)
Prints and photographs	13,262	3,914	1,845
Historical audiovisuals	4	430	792
Processing:			
Books cataloged	4,298	4,675	4,328
Modern manuscripts cataloged	243 (in ft)	1,437 (in ft)	2,018 (in ft)
Pictures cataloged	34,745	31,605	14,751
Citations indexed	5,134	now being done by Index Section	now being done by Index Section
Public Services:			
Reference questions answered	22,506	22,414	24,034
Onsite requests filled	13,183	13,439	14,843

SPECIALIZED INFORMATION SERVICES

Dr. Steven Phillips
Associate Director

The National Library of Medicine (NLM) Division of Specialized Information Services (SIS), offers a broad array 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).

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established more than 40 years ago within the NLM 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 technologies to provide more rapid and effective access to toxicological and environmental health information for a wider audience. 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 a 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 climate change and human health and nanotechnology.

SIS has been involved in the development of HIV/AIDS information resources for over 20 years, and we now have responsibility for several collaborative efforts in information resource development including a major outreach effort to the organizations serving those affected by HIV/AIDS.

The Division's outreach efforts 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 has significantly expanded its activities. In addition to continuing its research and development efforts as part of the Bethesda Hospitals' Emergency Preparedness Partnership (BHEPP), DIMRC has responded to several international crises, including the earthquake in Haiti, and expanded and improved the information resources available from NLM in this topic area.

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. This home page was simplified and a single search box added to improve access to the various programs and resources. 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

The Toxicology and Environmental Health Information Program (TEHIP) has evolved from its early days creating a toxicology database to an integrated service of toxicology information resource development and management, using the newest technologies to effectively reach a widening audience of researchers, scientists, health workers, and the general public. Central to this has been the Internet and the prevalence today of Web resources, and thus our TEHIP Web home page offers many approaches to a wide array of resources, including Enviro-Health Links that can be created quickly to respond to a current event or issue, such as the Gulf of Mexico oil spill (http://sis.nlm.nih.gov/enviro.html) links to our TOXNET cluster of databases are also provided on the page, allowing users to do a quick search of all resources or go to TOXNET for more specific searching.

TOXNET (TOXicology Data NETwork) is a cluster of databases covering toxicology, hazardous chemicals, environmental health and related areas. Some of these were started in the early days of days of database creation in the Division. Others are new to TOXNET, providing access to resources created at other sites. Enhancements to TOXNET are continuing, based on user feedback/requests and routine upgrades/additions of data and capabilities. Databases in TOXNET include:

Drugs and Lactation (LactMed), which provides information on drugs and other chemicals to which breastfeeding mothers may be exposed. It includes information on the levels of such substances in breast milk and infant blood, and the possible adverse effects in the nursing infant and includes links to other NLM databases. An app for using this valuable resource on mobile devices is under development.

Hazardous Substances Data Bank (HSDB), a peer reviewed database focusing on the toxicology of approximately 5,500 potentially hazardous chemicals. This flagship database was enhanced in FY2010 with records on non-materials and the chemicals involved in the Gulf oil spill clean up. We continued planning for a re-engineering of the HSDB to better incorporate new resources important to today's toxicologists.

Integrated Risk Information System (IRIS), a database from the US Environmental Protection Agency (EPA) containing carcinogenic and non-carcinogenic health risk information on over 542 chemicals.

International Toxicity Estimates for Risk (ITER), a database containing data in support of human health risk assessments. It is compiled by Toxicology Excellence for Risk Assessment (TERA) and contains over 680 chemical records. These side-by-side comparisons of international risk assessments include links to source documentation.

Chemical Carcinogenesis Research Information System (CCRIS), a scientifically evaluated and fully referenced data bank, developed by the National Cancer Institute (NCI) and until now maintained by SIS, with over 9,500 chemical records with carcinogenicity, mutagenicity, tumor promotion, and tumor inhibition test results. We will not have funding to continue this in FY2011.

Genetic Toxicology (GENE-TOX), a legacy toxicology database created by the US Environmental Protection Agency (EPA) containing genetic toxicology test results on over 3,000 chemicals. New data is not currently being added.

TOXLINE, a bibliographic database providing comprehensive coverage of the biochemical, pharmacological, physiological, and toxicological effects of drugs and other chemicals from 1965 to the present. TOXLINE contains over 4.3 million citations, almost all with abstracts and/or index terms and CAS Registry Numbers.

Development and Reproductive Toxicology/ Environmental Teratology Information Center (DART/ETIC), a bibliographic database covering literature on reproductive and developmental toxicology. This database is no longer funded by the multi-agency group that created it, but it is still searchable as a distinct entity.

Toxics Release Inventory (TRI), a series of databases that describe the releases of toxic chemicals into the environment annually for the 1987-2009 reporting years. TOXMAP provides an alternate entry into the TRI data.

ChemIDplus, a database providing access to structure and nomenclature authority databases used for the identification of chemical substances cited in NLM databases. ChemIDplus contains over 380,000 chemical records, of which over 295,000 include chemical structures. ChemIDplus includes some toxicity data as well as locators to many important national and international listings of chemicals.

Household Products Database, which provides information on the potential health effects of chemicals contained in more than 10,000 common household products used inside and around the home. New categories of products were added this year.

Haz-Map, an occupational toxicology database designed primarily for health and safety professionals, but also for consumers seeking information about the health effects of exposure to chemicals and biologicals at work. 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.

ALTBIB, a bibliographic database on alternatives to the use of live vertebrates in biomedical research and testing, developed as part of NLM's participation in the Interagency Coordinating Committee to Validate Alternate Methods. Work began this year to update the Web access and content of the ALTBIB resource.

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

TOXMAP, a Geographic Information System (GIS) system that uses maps of the United States to help users visually view data about chemicals released into the environment and easily connect to related environmental health information. Enhancements released in FY2010 include updated mortality data and EPA Superfund NPL data and new links to AirNow and NCEA's HERO. This year we initiated work on a "next generation" TOXMAP using ESRI ArcGIS server software.

Tox Town was enhanced with new content (in English and Spanish) in the neighborhoods of Tox Town, Tox City, Tox Farm, Tox Port and a US Mexico Border scene. Additional chemicals were added, such as bisphenol A. Tox Town was demonstrated at several educational conferences and promoted for use by teachers. Tox Town now features changing highlights on its home page on new resources and special events related to environmental health, such as National Poison Prevention Week.

ToxMystery, an interactive Web site for children between the ages of 7-10, was released at the end of FY2006. It provides an animated game-like interface, which includes

finding potential chemical hazards in a home and includes fun sound effects.

Drug Information Portal, providing current drug information for over 150,000 drugs with links to many credible additional online resources. During FY2010, changes were made to the search interface to allow easier selection of drug name and searching by category of drug. At the end of FY2009, we released Pillbox, a new drug information resource that focuses on pill images as well as drug names and other physical characteristics information. This year we continued to work with the FDA on the development of this resource, which is an integral part of the patient-safety initiative with the FDA.

Dietary Supplements Labels Database, developed with the NIH Office of Dietary Supplements (ODS), and enhanced with additional labels from the many brands available in the marketplace. This will continue as the ODS prototype label database intended primarily for researchers is developed.

The Disaster Information Management Research Center

The SIS Office of the Disaster Information Research Center (DIMRC) coordinates disaster health information management activities across NLM. In FY2010, the DIMRC office continued to strengthen earlier efforts to identify and provide access to disaster health information resources and to develop prototypes for informatics solutions to hospital emergency response activities. Major initiatives in FY2010 were the development of Web pages and other resources in response to several major disasters including the Haiti earthquake and cholera outbreak and the Gulf of Mexico oil spill. Web pages for each of these events quickly were developed in response to each of these events. NLM was the first federal agency to launch a Web page addressing the health effects of the oil spill.

Tools for Emergency Response

NLM launched its first apps for the iPhone/iPod touch/iPad in FY2010 with the Wireless Information System for Emergency Responders (WISER) and the Radiation Emergency Medical Management (REMM) resources now available from the iTunes store. WISER is a system designed to assist emergency responders in hazardous materials incidents and contains information on chemical, biological, and radiological agents. WISER is available on the Web and to download for Windows, Windows Mobile, iPhone/iPod touch, Blackberry, and Palm OS. REMM, a Web and mobile-device tool provides guidance to health care providers about the clinical diagnosis and treatment of radiological/nuclear events. REMM is available on the Web and can be downloaded for Windows, iPhone, iPod touch, Blackberry, Windows Mobile and Palm OS. In addition, DIMRC continued work on the Chemical Emergency Medical Management (CHEMM) tool. CHEMM is expected to be publicly released in 2011 and will provide just-in-time medical management guidance to health professionals as well as assistance with planning and training for mass casualty chemical events. CHEMM will include tools for modeling exposure to chemicals and provide syndromic treatment guidance. SIS continues to work with the HHS Office of the Assistant Secretary for Preparedness and Response (ASPR) on both REMM and CHEMM.

Disaster Information and Outreach

In FY2010, much of the DIMRC Web site redesign was completed with a public release date anticipated by the end of October. The new site will have a fresh look and feel and enhanced information including an RSS feed of new references to disaster medicine and public health literature, new topic pages, and easier navigation. Major enhancements were made to the Resource Guide for Public Health Preparedness, a grey literature resource to disaster health information to improve searching, navigation, and content. The Disaster Information Specialist pilot program continued at several institutions including Sarasota Memorial Hospital, New York University, and HHS/ASPR. Each institution continued to find unique ways in which librarians can assist emergency personnel from serving as informationists for researchers, health care professionals, and policy makers, to providing new resources and training, and serving on disaster preparedness committees at the local or regional level. The Disaster Outreach Listserv continued to grow and reached over 450 subscribers in FY2010. Plans for a Disaster Information Outreach Symposium began this year. The symposium will be March 29-30, 2011. DIMRC also began working with the Medical Library Association (MLA) on the development of a Disaster Information Specialization continuing education program to be offered by MLA. Several hundred additional disaster organizations were added to the Directory of Information Resources Online (DIRLINE) and a widget for searching these organizations is under development. DIMRC also continued to work on the Central American Network for Disaster Health Information (CANDHI) with the Pan American Health Organization/World Health Organization (PAHO/WHO) and the Regional Disaster Information Center for Latin America and the Caribbean (CRID). DIMRC also began working with PAHO and CRID on the development of a Haiti Disaster Information Center to ensure that vital health information is not lost.

Research and Development

The DIMRC office continued to coordinate and conduct a number of informatics and communications research and development and pilot 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 staff worked on several projects including development of a back-up digital communication system via the Military-Affiliated Radio System, and the development of a virtual world prototype for training hospital personnel in the nationally mandated Hospital Incident Command System. Several early-stage prototypes were tested at the Collaborative Multi-Agency Exercise (CMAX) in October 2009. Information gathered at the exercise provided guidance for further development of the projects in FY2010. Further testing will occur at the next CMAX in October 2010.

Partnerships

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

SIS Outreach Initiatives

United Negro College Fund Special Programs/NLM – HBCU Access Project, one of NLM's major outreach projects with Historically Black Colleges and Universities (HBCU), continued during FY2010 and awarded four HBCUs small grants to develop and implement projects that help to increase the awareness and utilization of NLM resources on campuses and in their communities. The annual June workshop had as its theme "Capacity Building for e-Health at HBCUs" and featured keynote addresses by Phyllis Champion, Executive Director of the Association of Minority Health Professions Schools and Betsy Humphreys, NLM Deputy Director. This was the ninth year for the two-day-workshop held at NLM Lister Hill Center. In addition, on the following day, an NLM online database training session was conducted at the University of the District of Columbia (UDC), Washington, DC, for about 30 conference attendees with a fourth day of technical assistance in developing and writing proposals held.

The mission of the Environmental Health Information Partnership (EnHIP) is to enhance the capacity of minority serving academic institutions to reduce health disparities through the access, use and delivery of environmental health information on their campuses and in their communities. The March 2010 meeting, held at NLM, featured Climate Change and Human Health as its theme. Speakers at that meeting included Ralph J. Cicerone, PhD, President, National Academy of Sciences; John Balbus, MD, MPH, Senior Advisor for Public Health, National Institute of Environmental Health Sciences; and Daniel Wildcat, PhD, Haskell Indian Nations University. Dr. Wildcat presented Dr. Lindberg with a copy of his book, Red Alert! Saving the Planet with Indigenous Knowledge. The EnHIOP meeting included representation from 14 HBCUs, 3 tribal colleges and 3 Hispanic-serving institutions.

Chickasaw Health Information Center (CHIC) is a project that was initiated by the Sacred Root Tribal Information Fellows from the Chickasaw Nation and is a partnership among the Chickasaw Nation, NLM, and ComputerCraft, a technology company owned by members of the Chickasaw Nation. The CHIC was originally located in the Carl Albert Indian Health Facility in Ada, Oklahoma, and equipped with two workstations and a printer along with brochures and other NLM and CHIC materials. It was staffed full-time by a trained tribal member. The Nation recently opened a larger health facility and moved the CHIC to a central location in the main lobby of the building.

Outreach to Minority Health Professionals

NLM's efforts to increase both the awareness and use of NLM online resources by minority health professionals continued last year. Hands on training classes or presentations and demonstrations were offered at the National Medical Association's (NMA) Annual Convention and Scientific Assembly, the National Black Nurses Association's Annual Institute and Conference, the Student National Medical Association's Annual Medical Education Conference and at several NMA regional meetings. In addition to these general training programs, SIS staff participated in NMA's Project I.M.P.A.C.T. clinical trials community forum held in Durham, NC. Project I.M.P.A.C.T. educated both physicians and the public about the value of clinical trials.

In 2010 SIS staff initiated outreach efforts in partnership with Historically Black Greek Letter Organizations. These fraternities and sororities are very active in community outreach and philanthropy. staff worked with Kappa Alpha Psi Fraternity and Alpha Kappa Alpha Sorority, in areas of health literacy and health awareness.

Outreach to Librarians

NLM staff worked with the Black Caucus of the American Library Association to coordinate and present a health literacy track as part of their 7th National Conference of African American Librarians in Birmingham, Alabama.

NLM continues to support the training and education of minority librarians through the 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.

Special Population Web Sites

Women's Health Resources (WHR), a collaborative with the Office of Research of Women's Health, part of the Office of the Director at NIH, was featured on page one as a "Research Highlights" and page 112 in the "Highlights of NIH Women's Health and Sex Differences Research 1990-2010" publication (PDF,1.6 MB) (http://orwh.od.nih.gov/ORWH_Highlights_2010_508.pdf) This publication was compiled to celebrate the 20th Anniversary of the Office of Research on Women's Health at NIH.

Social media tools are being utilized for promotional awareness of WHR. In April a Facebook and Twitter account were created for WHR: Twitter: http://twitter.com/WomensHealthNIH, Facebook: http://www.facebook.com/WomensHealthResources. WHR currently has over 3,000 Twitter followers and 290 Facebook "likes".

The American Indian Health Web Portal (AIH) User Group expanded with four additional members from geographic areas not previously represented in the group as well as an information specialist to help ensure that new information that becomes available is added to the Web portal. Much emphasis was placed on expanding the Our Stories section, audio narratives related to health and diseases. As a result of an associate project, fourty such items were added to the Web site.

The Arctic Health Web Portal launched a new section for "Climate Change" (http://www.arctichealth.org/climatechange.php). This section has over 1,000 links to reports, videos and Web sites. Categories in the section include: events, news, effects of climate change on human health, resilience and adaptation in Northern communities, traditional knowledge and witnesses to change, changing ecosystems, ice and climate, understanding climate change, artists' perspective and climate change organizations.

SIS is a partner in the Refugee Health Information Network (RHIN), which is 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, multi-cultural 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. The emphasis this year has been on the issue of obtaining materials for new refugee populations entering the US.

AIDS Resources and Activities

NLM is the project manager for the multi-agency AIDSinfo service (aidsinfo.nih.gov) and its Spanish language companion site, infoSIDA. This service provides access to federal HIV/AIDS treatment guidelines, AIDS-related clinical trials information (through http://clinicaltrials.gov/), and prevention and research information. The American Customer Satisfaction Index (ACSI) has been running on the infoSIDA site, and work is underway to include it on AIDSinfo as well. In 2010, at the request of the NIH Office of AIDS Research, AIDSinfo significantly increased the level of support provided to the panels that develop the Guidelines for the Prevention and Treatment of Opportunistic Infections (both adult and pediatric), Guidelines for the Use of Antiretroviral Agents in Pediatric HIV Infection, and the Recommendations for the Use of Antiretroviral Drugs in Pregnant Women (perinatal transmission). The staff continues to provide support for the Use of Antiretroviral Agents in HIV Infected Adults and Adolescents panel. As detailed in the Evaluation section, usability studies have been conducted on AIDSinfo and infoSIDA and the outcomes will be used for the redesign of the Web site.

The National Library of Medicine (NLM) has continued its HIV/AIDS-related outreach efforts to community-based organizations, patient advocacy groups, faith-based organizations, departments of health, and libraries. This program, which started in 1994, provides support to design local programs for improving information access for AIDS patients and the affected community as well as their caregivers. Emphasis is on providing information or access in a way meaningful to the target community. Projects must involve one or more of the following information access categories: information retrieval, skills development, Internet access, resource development, and document access. In FY2010, NLM made awards to the following organizations: AIDS Educational Global Education System (AEGiS), Metro TeenAIDS/Washington Area Consortium on HIV in Youth, Philadelphia FIGHT, P.I. Advocates International (PIA), Renz Addiction Counseling Center, University of Kentucky Research Foundation, University of New Mexico Health Sciences Center, Urban League of Flint.

HIV/AIDS Training

Equal Access Initiative: HIV/AIDS Information Resources from the National Library of Medicine

The NIH Office of AIDS Research (OAR) and the National Minority AIDS Council (NMAC) partnered with NLM starting in 2007 to expand the scope of the Equal Access Initiative Computer Grants Program by including a series of training classes for recipients of the computer grants. NLM staff conducted two 4-hour training sessions for 129 recipients of the Equal Access Initiative (EAI) Computer Grants Program during the USCA in Orlando, Florida. Participants received training on NLM resources, a train-the-trainer tool kit which included NLM promotional materials, a "Quick Reference Guide to HIV/AIDS Treatment Resources" CD, and a workbook. Resources from the training program can be found online – http://www.sis.nlm.nih.gov/outreach/equalaccess.html. A

program evaluation showed a significant increase in the number of participants rating their ability as excellent after the training relative to how they perceived their ability before the training in perceived ability to be able to locate reliable HIV/AIDS information resources online and find evidence-based research articles in PubMed.

SIS K-12 Initiative

In FY2010, SIS significantly expanded its K-12 program, which includes resource development, program development, outreach, and research components. Two new resources were developed: 1) Environmental Health Student Portal, a Web site for middle school students and teachers and 2) a health education book that supports middle school students dealing with parental addition to food, alcohol, and tobacco. The Portal is publicly accessible on http://kidsenvirohealth.nlm.nih.gov/ ; the book is in the final stages before being released in electronically on our web site in pdf format. Both of these resources were developed based upon feedback and review at every step. The middle school environmental health portal was developed with significant input from middle school teachers and several rounds of focus testing and dialog with science teachers. The health education book, written by a published author of children's books, was reviewed by teachers and several NIH institutes and offices.

In the area of program development, we initiated collaboration with the American University School of Education and Cabin John Middle School in Montgomery County, Maryland, with the goal to develop lesson plans to be located on the middle school environmental health portal. The lesson plans were developed within the context of an environmental health afterschool club. The collaboration and the work will continue in FY2011. Research activities pertaining to our K-12 program are described in the "Research Activities" section of this report.

Our outreach activities included a number of presentations to middle school teachers and school administrators. NLM presented information on its K-12 environmental health resources to science teachers at the annual Maryland Association of Science Teachers conference in October 2010. In March 2011, NLM presented with the National Institutes of Health/NIGMS at the National Science Teachers Association annual conference to give a more global introduction to all NIH K-12 resources. Finally, to kick off National Lab Day, teachers from the Maryland Association for Science Teachers were invited to the National Library of Medicine for a hands-on workshops and presentations on all NLM K-12 resources.

Evaluation Activities

In FY2010, we received competitive NIH Evaluation set-aside funding for four evaluation studies. Three awards were to conduct usability evaluations, informing scheduled upcoming redesign projects. The first award was to evaluate AIDSinfo, a comprehensive site with information on AIDS / HIV treatment, prevention and research, and its Spanish language counterpart, infoSIDA. The evaluation was completed, the findings were communicated to the design team, and redesign work has commenced. The second award was to evaluate the new design of Radiation Emergency Medical Management Web site, which provides evidence-based data for healthcare professionals about radiation emergencies. The evaluation is completed. The third award was to evaluate TOXNET, a collection of toxicology, environmental health, and chemical databases. This evaluation is currently in progress. The fourth award from the Evaluation Office was to develop an evaluation plan for NLM AIDS Community Information Outreach Program. SIS conducted an extensive search for a contractor with in-depth theoretical and practical expertise in evaluating the impact of health information outreach. We are currently reviewing two proposals from academic institutions, and plan to start the evaluation in the spring of 2011. Finally, another SIS evaluation activity commenced in FY2010 involved employing Foresee Results to conduct ACSI web survey on AIDSinfo and infoSIDA Web sites.

Research Activities

SIS participated in a number of research activities, pertaining to three broad areas. The first research pertains to evaluating methods and strategies of health information outreach. We completed a research paper, based on our 2009 study, which aimed to adapt Web 1.0 evaluation methods to virtual worlds. The study was contextualized in SIS application in Second Life virtual world, Tox Town in Second Life. In our other evaluation-related study, we have been working on a review of published studies of community health information outreach efforts. The goals are to identify characteristics of successful health information outreach programs, as well as effective evaluation practices for capturing best practices. The study is ongoing. The second research area is consumer health informatics. We collaborate with a team from the University of Wisconsin, Madison and Drexel University on a study of factors that affect patients' comprehension of consumer health materials. Our FY2010 work on the project involved statistical data analysis and work on two manuscripts, based on the project. The third research area involves K-12 teaching and learning in the area of health and science. This research is part of our comprehensive K-12 resource development and outreach program. An SIS representative delivered an invited keynote lecture at international conferences, "Environment and Health in Science Education" in Zurich, Switzerland. We have also established collaboration with the American University

School of Education, studying how teachers design environmental health activities to support the development of various competencies. Finally, we worked on an invited manuscript summarizing our research program and approach to health education in a science classroom (the manuscript now have been submitted).

In FY2010, SIS awarded a contract for a Pillbox solid dosage medication imaging research and development project. The FDA funded through an interagency agreement with FDA. One thousand unique medications were acquired through a partnership with Florida A&M University's College of Pharmacy, with the assistance of Dr. Henry Lewis, III. The project team developed a standardized methodology to create high-resolution images of medications suitable for applications in Health IT, as well as image processing and informatics research. Documentation produced as a result of this project include a standard for product images, which NLM has presented to FDA to be included as guidance to the pharmaceutical industry for all pharmaceutical images included in drug labels. Additionally, a methodology document provides a how-to guide for photographers to create images of this quality and overcome challenges specific to macro photography of this type. With this work, NLM is supporting FDA in eventual changes to the drug labeling regulations. The images produced through this project are now included in the Pillbox Web site and Web service.

Social Media/New Media Activities

SIS established a Social Media Working Group to explore how the division can harness the power of social media in support of an integrated health communications approach for both internal communication among staff as well as with external users. The group is exploring policies, standards, and best practices in order to guide SIS as we embark on using social media for outreach. Experiments led by this group included the use of micorblogging software such as Yammer and setting up an internal Facebook account. This group also provided advice to project managers as new Facebook and Twitter accounts are being planned for various programs.

SIS developed or significantly expanded several smartphone "apps" including Health Hotllines, AIDSinfo glossary in English and Spanish, WISER all for the iPhone. Development of Android and Blackberry versions are underway. A Mobile version of AIDSinfo was also released. SIS added several APIs to NLM's Web site: AIDSinfo drug database, ChemSpell (chemical spell-checking), DIRLINE, Pillbox beta, and TOXNET.

The widget for disaster-related organizations in DIRLINE was finalized and is now on the main search page of DIRLINE. This widget limits a DIRLINE search to agencies within a state or all agencies tagged "DIMRC."

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 NLM. The Center 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.

Lister Hill Center research staff is drawn from a variety of disciplines including medicine, computer science, library and information science, linguistics, engineering, and education. Research projects are generally conducted by teams of individuals of varying backgrounds and often involve collaboration with other divisions of the NLM, other institutes at the NIH, other organizations within the Department of Health and Human Services, and academic and industry partners. staff members regularly publish their research results in the medical informatics, computer and information sciences, and engineering communities. The Center is visited by researchers from around the world.

The LHNCBC is organized into five major components: Cognitive Science Branch (CgSB), Communications Engineering Branch (CEB), Computer Science Branch (CSB), Audiovisual Program Development Branch (APDB), and the Office of High Performance Computing and Communications (OHPCC).

An external Board of Scientific Counselors meets semi-annually to review the Center's research projects and priorities. The most current information about the Lister Hill Center research activities can be found at http://lhncbc.nlm.nih.gov/.

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

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

NLM Personal Health Record

The goal of the NLM Personal Health Record (PHR) project is to help individuals manage the health care of either relatives or themselves. The PHR serves as a testbed for validating and improving NLM clinical vocabularies, studying consumers' use of PHR systems, studying the potential of PHR-based educational reminder systems to improve prevention, and as a potential vehicle for gathering patient information during clinical trials.

The NLM PHR supports the entry and tracking of key measurements and test results, prescriptions, problems, and immunizations, future health appointments. It will produce digital and paper copies of its contents in various formats. Users can get access to MedlinePlus information resources by clicking the icon adjacent to the name of any prescription drug, medical condition, or surgery that they enter into the system. The PHR automatically assigns codes to the medications, observations, and problems as users enter them and utilizes NLM-supported terminologies that HHS recognizes as national standards. By using vocabulary standards and coding, the PHR can "understand" its content and provide numerous benefits such as computer-generated personalized reminders about preventive care or healthy behaviors and automatic calculations based on other values in the PHR (e.g. calculate body mass index based on height and weight entered in the same panel). The PHR provides direct links from patient-entered conditions, drugs, or surgeries to MedlinePlus and other respected consumer information resources. In FY2010, researchers continued to expand and improve the capabilities of the PHR. Developers gave users the ability to set the number of days before their medications expire, to initiate a mail order refill, or to obtain follow-up testing. The PHR provides easy ways to generate input forms for any kind of health information that the patient wants to track

Researchers reviewed and enhanced the controlled vocabulary for more than 2,000 condition names and synonyms and more than 300 surgery procedure names and synonyms by enriching the synonymy, providing the consumer-friendly name when feasible, and adding SNOMED codes, when available, to these items.

Developers completed a major revision of the decision rules authoring system. In the first version, authors needed to use a syntax that required programming skill. In the latest version, users pick from a series of menus and input fields on a Web form. To build a base rule, the author identifies the PHR table of interest, e.g. medications, conditions, allergies, etc., and then identifies the subject and the predicate of the rule. A rule author can identify a subject of the rule by name, or class. For example, if the users picked medications as their table of

interest, they could identify the subject by name, e.g. Simvastatin oral, or by class, e.g. cholesterol lowering drug, and then add criteria such as status = active, or date started was before January 1, 2009. Users could define another rule to determine whether a patient's last LDL cholesterol was greater than 130 mg/dl during the last year. These two rules could be combined into a third rule for reminding PHR record holders to ask the doctor about the use of diet and/or a cholesterol lowering agent if their LDL cholesterol was elevated and their medication list did not contain any active cholesterol lowering agents. As research progresses, the rules will support increasing complexity. Researchers than completely rewrote all of the existing rules, and added a substantial number of new rules using this new system.

Developers added a number of other capabilities to the PHR including:

1. pop-up forms for capturing information from almost any questionnaire or laboratory panel. The only prerequisite is that the information that defines the questionnaire (or lab panel) must be in the PHR's master knowledge database and follow the LOINC form structure. Currently the PHR has access to more than 500 such "questionnaires."
2. a new data table driven rule system that supports reminder rules.
3. new tables for identifying medication content by ingredient and a system to warn PHR users when they enter a prescription that contains the same ingredient as a previously entered one.

The LHNCBC continues to work with NLM and others to ensure that policies concerning the PHR are developed and in place. This young project addresses the longstanding NLM interest of facilitating health care management and is closely aligned with the NLM strategic plan. It will help refine the message and vocabulary standards that NLM has supported and also provide another consumer entry point to a rich trove of patient-oriented data.

Use of Surescripts Prescription Data in Direct Patient Care

An important part of patient assessment in the Emergency Department (ED), medication history can have significant impact on the diagnosis and treatment of a patient's problems. However, manually-acquired medication histories are prone to inaccuracies. An electronic prescription history, which can be obtained through sources like Surescripts, a consortium of major Pharmacy Benefit Managers (PBM) and the largest e-prescribing network in the U.S., has data that could improve these histories. Surescripts handles over two billion prescriptions per year and can provide prescription benefit and history information for an estimated 65 percent of all prescriptions paid for by private insurance in the United States.

Partially-funded by the Bethesda Hospitals Emergency Preparedness Partnership (National Naval Medical Center, NIH Clinical Center, and Suburban Hospital), this project established a connection between Suburban Hospital and Surescripts to allow real-time retrieval of a patient's prescription history, and studied the feasibility and value of using prescription history information from Surescripts in direct patient care. This external source of medication information that can be obtained automatically could be both time- and life-saving in disaster circumstances, when the normal route of obtaining patients' medication history is likely to be disrupted due to over-strained medical staff or special patient circumstances (e.g. unconscious patients).

As a quality assurance process, we compared the history obtained by the ED nurse to Surescripts information for all ED patients for three months. We found that neither source was complete by itself (nearly 20 percent of all potentially current medications were not captured in the manual medication history and 25 percent of the medications were missing from Surescripts data); however, Surescripts does add substantial information to the medication histories collected by the ED personnel.

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

Suburban Hospital changed their HL7 engine in 2010, which affected all HL7 transactions, including the feed that we use to exchange data with Surescripts as well as to collect de-identified research data to study the value of Surescripts' medication history reports. The developers completely redesigned the HL7 interface, the associated relational database schema, and the hardware to represent and capture the new data stream seamlessly.

Surescripts transmits a raw medication history report for each patient that lists every prescription and refill, each on a separate line, that providers find difficult to review and digest. We have developed a pilot report that we believe will be easier to comprehend. In our new approach to reporting, the raw Surescripts data is reprocessed, clustered by medication names, sorted by date, associated with dispensed drug duration information (when available), and at the end, a one-year-prescription history of the patient is plotted on a timeline graph. In the latest iteration of our design, we can also convey pharmacy

and prescriber information (when available) and plot each drug dispensation separately so that care providers, with the help of their patients, can distinguish the actual data points from the noise.

EMR Database Research and Development

LHNCBC developed a general purpose longitudinal database structure to investigate secondary use of data collected in electronic medical records and shared this structure with the owners of the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) database as well as the AMIA-NLP working group. Both groups will use our schema for design and development of their repositories of clinical data.

This year LHNCBC acquired a new update of the MIMIC II clinical data. We use this de-identified dataset under a restricted-use Memorandum of Understanding (MOU) to populate our database. The current version of this MIMIC II dataset consists of clinical data for over 32,000 ICU encounters and more than 26,000 patients. It addition to over 11,000,000 laboratory results and over 400,000 clinical notes, the dataset now contains information on a patient's death at any time after the patient's discharge. Altogether, the dataset carries over 200,000,000 discrete observations. The current version of the data contains the Simplified Acute Physiology Score (SAPS) designed to measure the severity of disease and predict mortality for patients admitted to intensive care units. The initial scores were developed by the MIT research team. In the process of replicating the scores, we found additional variables that could be used to compute the scores for almost all patients and reconciled discrepancies with the MIT team. We also determined that some of the extreme values for the variables contributing to SAPS are correct for a given patient and others are device or data entry errors. We are developing methods for quality assurance of data using structured information and clinical notes.

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

In FY2010, the LHNCBC worked with the National Cancer Institute (NCI) and the American Society for Colposcopy and Cervical Pathology (ASCCP) to put one of our imaging programs, the Teaching Tool, into operational use for the assessment of professional knowledge and skills in the field of colposcopy. The Teaching Tool is being used by 52 resident programs in Ob/Gyn and Family Practice, and over 200 individual online exams have been given. In March 2010 the ASCCP acknowledged our work by giving the organization's Award of Merit to two members of the Teaching Tool development team.

NCI deployed another of our imaging programs, the Boundary Marking Tool, at sites including the University of Oklahoma Health Sciences Center and sites in Guanacaste, Costa Rica, and Irun, Nigeria to collect and annotate colposcopy images for the creation of a worldwide database for cervix research. We also began providing public access to the Boundary Marking Tool code under a modified Berkeley open source license, through a link on the CEB Web page. We continued development on both of these tools to improve functionality and usability; part of this effort included carrying out a formal usability study for the Boundary Marking Tool with participants at the biennial meeting of the ASCCP in Las Vegas in March 2010. We also have continued development of the Multimedia Database Tool by adding several thousand graphical annotations of cervix images to the two databases accessible through this tool. staff are currently working with NCI to add pH measurement data to these databases to support NCI study of the correlation between vaginal microflora and HPV infection. We have begun collaboration with a new group of NCI-sponsored researchers at the University of New Mexico and are working toward expanding the capabilities of the CEB Virtual Microscope tool to support a study of histology diagnoses of the cervix in glass slides versus digital images. If this study indicates positive results for digital microscopy research in this field, we anticipate supporting further NCI histology studies with this tool. Also this year, to benefit our NCI collaborators, we created a new graphing capability to provide a compact visual representation of the spatial distribution of pathology across the surface of the cervix.

In September 2010, along with representatives of NCI and the ASCCP, we presented our work with the Teaching Tool to the NLM Board of Regents.

Content Based Image Retrieval (CBIR)

CBIR is an active research area at the Lister Hill Center with several objectives. One is to develop techniques to introduce automation into our existing cancer research tools. For example, our CervigramFinder automatically indexes and allows retrieval of cervigrams using shape, color and texture features. This system therefore contains the key elements needed to augment the Boundary Marking Tool with an automated assist for the user in marking boundaries of regions of medical significance. Researchers evaluated the CervigramFinder for usability and acceptance at the biannual meeting of the American Society for Colposcopy and Cervical Pathology (ASCCP) in 2010, and developers are using the evaluation results are

to improve the tool. CBIR research is supported by the Cervigram Segmentation Tool that combines several image segmentation and shape similarity algorithms. This tool enables researchers to select the optimal algorithms that automatically segment regions for image indexing. We also developed a third tool, MOSES, as a service for the evaluation of segmentations done by multiple experts, to reduce the inter- and intra-observer variability in marking boundaries of significant regions in cervigrams. Together, these three tools form a unified system for cervigram image analysis.

In addition, we developed a hybrid CBIR system for x-rays in the NHANES II collection by linking two geographically separated CBIR systems (SPIRS at the Lister Hill Center, and IRMA at the University of Aachen in Germany) to exploit the characteristics of each. SPIRS/IRMA allows a user to retrieve detailed medical image data in large collections of heterogeneous images of varying modalities, presentations and anatomy. For example, a user may narrow a search to *spinal x-rays* (using IRMA) and then retrieve specific x-rays containing *osteophytes* through SPIRS. In FY2010, we extended the capability of SPIRS with research into improved shape matching methods, resulting in more efficient shape matching as well as a better understanding of the optimal number of shape features needed. We implemented a new feature (distance across shape) and used a Support Vector Machine (SVM) classifier to reduce the *semantic gap* between diagnostic labels used to describe the pathology on the vertebrae and the shape characteristics extracted by feature extraction algorithms.

CBIR has been used to index illustrations in medical journals by using image features, in combination with text processing of figure captions and in-document text mentions. This research is aimed at enriching the user experience of searching for relevant documents by including the contents of medical images, photographs, graphs and other illustrations found in articles. Techniques developed in this work were evaluated in the international ImageCLEF competition in 2010 and found to be successful. Over 15 image features were implemented and used in an SVM-based framework to detect modality (x-ray, ultrasound, CT, MR, etc.) and to compute similarity. Our efforts were ranked second among 12 teams from around the world, many of which were from industrial R&D labs. We also developed methods to describe images in a *bag-of-words* and a *bag-of-keypoints* representation analogous to those used in text-document retrieval. These were very successful in automatic coarse annotation of images. Research into improved automatic image annotation methods is ongoing.

We have explored the role of CBIR in extracting regions of interest (ROI) in images. One approach to identifying meaningful ROIs, and thereby annotating biomedical-article images, is by first extracting author-placed markups (or "pointers") such as arrows, asterisks, and alphanumeric characters. Novel methods were developed for each type of markup with over 87 percent accuracy in detecting arrows. Further research is ongoing**.**

CBIR has also been used in a new project for screening digital chest x-rays for pathology, such as tuberculosis and other pulmonary diseases, prevalent in third world countries. As an initial step in this project, we have developed image content analysis methods to automatically detect lungs and ribs in the x-ray images. Research into image feature extraction and machine learning methods for detecting and classifying images is ongoing.

Other avenues explored in this research area are distributed computing and use of Graphics Processing Units (GPU) for compute-intensive CBIR tasks, with a particular focus on image segmentation.

Research Toward Next Generation Scientific Publishing

Separate efforts are under way to explore techniques and technologies to create and use new and powerful forms of scientific publications that exploit the increasingly availability of various forms of multimedia. The goal of the first project, Interactive Publications Research (IPR), is to demonstrate a type of highly interactive multimedia-rich document that serves as a model for next-generation publishing in biomedicine. The research focuses on the standards, formats, authoring and reading tools necessary for the creation and use of such interactive publications (IP) containing media objects relevant to the biomedical literature: text, video, audio, bitmapped images, interactive tables and graphs, and clinical DICOM images such as x-rays, CT, MRI, and ultrasound.

The first tool LHNCBC engineers developed, *Panorama*, is for viewing and analyzing video, DICOM clinical images, tables, graphs and animations. Panorama, written in Java, was one of nine semi-finalists out of 70 entrants in Elsevier's Grand Challenge contest a year ago. In FY2010 we developed *Forge, a* tool for authors or publishers to create interactive publications. The Forge environment replicates the Panorama look and feel with the additional ability to edit and define the interactivity of each PDF annotation. Panorama is for viewing only and Forge is for both viewing and editing. These tools were recently demonstrated at the 2010 AMIA symposium and described in a publication (Thoma GR, et al. Interactive Publication: the Document as a Research Tool. *Web Semantics: Science, Services and Agents on the WWW*, 2010).

To support the future long term development of these Java tools in an open source environment, developers re-wrote the core code of Panorama and Forge to accommodate Eclipse plug-ins (http://marketplace.eclipse.org/) and thereby extend the functionality of these applications.

Taking advantage of our InfoBot system, we also enhanced the Panorama application to provide Annotation Concepts. This feature provides additional meaning and context to the article text. The text in an IP is sent to an

NLM servlet for processing to identify Unified Medical Language system (UMLS) concepts. An XML file is returned to Panorama that is parsed and provides linkouts for MedlinePlus, eMedline, Family Doctor, and also provides the preferred UMLS term and semantic group. Further work is ongoing to group concepts by semantic relationships, and other grouping ideas are being explored as well.

Engineers developed a prototype interactive publication and presented it to the National Center for Health Statistics (NCHS). This multimedia document contains Motion Charts using NHANES data from NCHS. Motion Charts is the Google widget version of Gapminder (created in Stockholm in 2005). NCHS is interested in creating IPs of their own but want to delay until they publish their conventional papers in the literature first.

We are also investigating several approaches for online access to IPs, since convenient and rapid access to these potentially large documents would be necessary in a practical deployment. In one approach, we modified Panorama to support Java's Web Start technology. Using this technology, standalone Java software applications can be deployed with a single click over the network. Java Web Start ensures that the most current version of the application will be deployed, as well as the correct version of the Java Runtime Environment (JRE). We are in discussion with a publisher to experiment with this and other approaches to provide access to IPs.

In a separate approach to new forms of publishing, LHNCBC is partnering with the Optical Society of America (OSA) to test the use of interactive publications in an operating journal. The goal is to evaluate software and database infrastructure that enables viewing and analysis of curated, supplemental biomedical source data published in conjunction with peer-reviewed manuscripts, evaluate the educational value of such an infrastructure, and explore the problems of archiving this medium. In order to accomplish these goals, OSA has published four electronic special issues of OSA journals on research topics which lend themselves to interactive publishing. Articles published in these special issues are peer reviewed and fully citable as OSA journal publications indexed in Medline. They are published on the Web in Acrobat format (PDF) with links to source data, videos, and other media objects. The links allow users to quickly and conveniently download these objects and visualize them using interactive viewing software designed to look like an Acrobat plug-in. The viewing software is freely available as a download for all computer platforms. The journal articles and data sets are open access and the source data and associated metadata are searchable and accessible from outside the publication. The project will also serve as an NLM testbed for archiving this new form of publication. The most recent special issue focusing on *Image in Diagnosis and Treatment of Lung Cancer*, published in April 2010, is available at http://www.opticsinfobase.org/oe/virtual_issue.cfm?vid=105.

Screening for Tuberculosis in Rural Africa

The goal of this new project is to develop a biomedical image processing system for clinically screening HIV positive patients in rural Kenya for lung disease with a special focus on tuberculosis, which is rampant. This project leverages the image processing, analysis, and communication expertise at the LHNCBC, and aligns with NIH and NLM policy and strategic planning objectives in global and rural health. LHNCBC initiated this collaborative project with Academic Model Providing Access to Healthcare (AMPATH), the largest AIDS treatment program in the world (more than 100,000 AIDs patients under treatment) and part of a large USAID collaboration.

Chest radiography is important to the detection and proper treatment of tuberculosis and other pulmonary infections which are prevalent among HIV-positive patients. As part of the collaboration, AMPATH is acquiring a lightweight FDA-approved digital x-ray imaging system that can be transported in existing SUVs and/or trucks to remote regions in Kenya at regular intervals in order to screen the high risk individuals and bring them into treatment programs. Initially, LHNCBC will obtain an existing de-identified collection of 1,000 film-based chest x-rays from Kenya that have been photographed using a digital camera. These images will be used to train image processing algorithms and machine classifiers to develop a screening system for relevant pathology.

Once the equipment is deployed, we will routinely collect data about the usage and durability of the equipment, the number of x-rays produced, and the performance of the screening algorithm to determine the value of this light weight approach and its long-term viability of such systems in other parts of the world.

Medical Article Records Groundtruth (MARG)

To exploit the images and labeled data collected routinely from the MARS operation for research, a validated test set for document image analysis was created for the computer science and informatics communities for research into advanced algorithms for data mining. The MARG database consists of images of journal articles, the corresponding OCR data, zones, labels and verified data obtained from the routine operation of the MARS production system.

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 Visitors Center, available anywhere through the Internet. 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.

Computational Photography Project for Pill Identification

As an exploratory program for content-based information retrieval to promote patient safety at a national level, the LHNCBC has undertaken a new project, *Computational Photography Project for Pill Identification* (C3PI), an authoritative, comprehensive, public digital image inventory of the nation's commercial prescription solid dose medications. Heretofore, we have conducted most of this work in partnership with the NLM Specialized Information Systems Division and the U.S. Veterans Administration to study content-based retrieval methods for medical image databases. NLM researchers have 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. Working with an expert team from Medicos Consultants, we are creating a collection of digital photographs of prescription tablets and capsules, creating high resolution digital photographs of the front and back surfaces of pharmacy samples, confirming that the images match the description of the medication, developing and matching the images of the samples to relevant metadata (including size descriptions, dimensions, color, and the provenance of the sample).

Virtual Microscope (VM) and Virtual Slides

LHNCBC has created an archive of virtual slides from the teaching set of glass slides from the Department of Pathology of the Uniformed Services University and other collaborating institutions. Researchers digitize, segment, and process each slide to simulate an examination of a glass slide under the microscope but with a Web browser. The collection preserves the specimen for posterity and allows viewing by users worldwide anytime. The system provides annotations and automatic linking to Medline/PubMed. A related collection of images from the AFIP fascicles allows users to search images and automatically link to Medline citations. In collaboration with the Massachusetts General Hospital, researchers are exploring the feasibility of a Virtual Slide mobile application and its use in training. Other collaborations will study its use in telemedicine and teleconsultation. The recent proliferation of iPhone and iPad devices required a modification of software to enable non-Flash capable devices to display virtual slides. Virtual slides are now accessible using both Flash capable and non-Flash capable computers and handhelds.

The Visible Human Project

The Visible Human Project image data sets are designed to serve as a common reference for the study of human anatomy, as a set of common public domain data for testing medical imaging algorithms, and as a testbed and model for the construction of image libraries that can be accessed through networks. The Visible Human data sets are available through a free license agreement with the NLM. They are distributed 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 3,200 licensees in 61 countries are applying the data sets 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 FY2010, we 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.

While designing a database of the parameters and variances defining the normal range of human anatomical structures and the dependencies and covariances between them, an attempt was made to glean the needed statistical data on bone size variation from the existing anatomical literature. Over 1,000 references were scanned. The literature was found to contain a description and images of the variation in shape that could be expected of each bone, but not the mathematical description of it variances.

3D Informatics

In FY2010, the 3D Informatics Program (TDI) expanded research efforts concerning problems encountered in the world of three-dimensional and higher-dimensional, time-varying imaging. The LHNCBC provides continuing support for image databases, continues to explore the growing need for image databases, including ongoing support for the National Online Volumetric Archive (NOVA), an archive a collection of volume image data. This collection contains 3D data from across medicine. The TDI group began migration of these data to the newly installed infrastructure based on the MIDAS software system, the same systems supported as part of an NLM project in Interactive Scientific Publication. Contributors to the collection include the Mayo Clinic Biomedical Imaging Resource and the Walter Reed Army Medical Center Radiology Department. The archive contains such integrated and multimodal data as virtual colonoscopy matched with recorded video from endoscopic interventions, time-varying 3D cardiac motion, and 4D MRI of a human hand. During 2009, the TDI group installed the necessary software and hardware infrastructure, including a Linux server and a MIDAS software system, to support interactive scientific publication at NLM. We continue to serve a broad community with these data, and seek to establish a leadership role through public data distribution.

We continued our collaboration with members of the LHNCBC APDB and the National Cancer Institute's

Laboratory for Cell Biology to visualize and analyze complex 3D volume data generated through dual beam (ion-abrasion electron microscopy) and cryo-electron tomography. This investigation centers on analysis of the spatial architecture of cell-cell contacts and distribution of HIV virions at virological synapses formed between mature dendritic cells and T cells. 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.

Collaborative work in telemedicine has increased the demand for large display technologies, but the cognitive pathways for understanding how the human visual system processes large-scale digital displays are not well understood. Previous TDI group achievements include the construction of high-end rendering systems for large displays, incorporating multiple GPUs in volume-rendering applications capable of rendering the full VHP Male dataset at real time rates. In 2010, we expanded our investigations in the question of presence and visualization, combining forces with the NCRR-funded Biotechnology Resource Center at the Utah Scientific Computing and Imaging (SCI) Institute and their access to the Magnetic Encephalography (MEG) equipment in the University of Utah, Department of Radiology. We also began a study of the difference between human-scale versus non-human scale perception, seeking changes in cortical processing of reading and eye-tracking from large wall displays versus hand-held book-sized displays. The result of such a study could have a far-reaching impact on how new digital user interfaces are developed in future generations.

In FY2010, the TDI study of the use of rapid prototyping technologies in Radiology generated considerable success. Our goal is to develop witness objects or phantoms, reference models as a public engineering and scientific standard for research in image-based computer-aided diagnosis. We have characterized the x-ray attenuation characteristics of some of the 3D-printing materials available at NIH. and are presently evaluating the use of contrast agents as printing materials to vary the appearance of the 3D models In FY2010, we successfully modified the 3D printing process through the use of contrast agents, primarily sodium iodide, to create 3D models that mimic human tissue when viewed with x-ray CT scanners. The goal is to create complex, anatomically-accurate models to test diagnostics systems and evaluate and compare their performance under known conditions. We were able to create models that correspond to CT scans of the Visible Human Project male dataset and demonstrate the possibilities for modeling soft tissue and metastatic disease. This work is conducted in partnership with the National Institute of Allergy and Infectious Diseases.

TDI also facilitated new collaborations and expanded research in medical imaging throughout 2010. With the appointment of a new postdoctoral fellow, OHPCC and the TDI Group began working on volume rendering of strain images from 3D ultrasound in a collaborative effort with the robotics group at Johns Hopkins University Department of Computer Science. This work was expanded over the summer with the participation of a Medical Informatics Training Program (MITP) graduate research fellow. New collaborations for OHPCC also included new work with the NIH Clinical Center's Center for Infectious Disease Imaging on a long-term study of pulmonary fibrosis, where a substantial cohort of affected patients and a corresponding control group contributed image data to enable research and development for computer aided diagnosis tools in pulmonary medicine. Findings on multimodal correlative medicine were developed and submitted for publication also with the participation of an NIH summer high school intern who is extending the project into an Intel National Science Talent Search project. Finally, OHPCC has published recently developed work on statistical deformation models (SDMs) for morphological image analysis based on Markov Random Fields.

Insight Tool Kit (ITK)

In FY2010, the Insight Toolkit celebrated a decade of service for biomedical image analysis with a symposium at the LHNCBC. The current official software release is ITK 3.20. Over 845,000 lines of openly available source code comprise ITK, making available a variety of image processing algorithms for computing segmentation and registration of high dimensional medical data on a variety of hardware platforms. ITK can be run on Windows, Macintosh, and Linux platforms, reaching across a broad scientific community that spans over 40 countries and more than 1500 active subscribers to the global software list-serve. A consortium of university and commercial groups, including OHPCC intramural research staff, provide support, development, and maintenance of the software.

ITK 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 Roadmap National Center for Biomedical Computing (NCBC), has adopted ITK and its software engineering practices as part of its engineering infrastructure. ITK also serves as the software foundation for the Image Guided Surgery Toolkit (IGSTK), a research and development program sponsored by the NIH National Institute for Biomedical Imaging and Bioengineering (NIBIB) and executed by Georgetown University's Imaging Science and Information Systems (ISIS) Center. IGSTK is pioneering an open API for integrating robotics, image-guidance, image analysis, and surgical intervention. International software packages that incorporate ITK include *Osirix*, an open-source diagnostic radiological

image viewing system available from a research partnership between UCLA and the University of Geneva and the Orfeo Toolbox (OTB) from the Centre Nationale D'Etudes Spatiales, the French National Space Administration. 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.

In FY2010, OHPCC and the ITK project ran two competitions to revise, accelerate, simplify, expand, and test the Insight Toolkit. Through the ITK Version 4 (ITK-v4) project, we are attempting to upgrade ITK for emerging computational platforms and meet research needs for the upcoming decade. The development of a major new version of ITK will help to continue our international leadership role in medical imaging research. A companion effort known as the 2010 ITK Algorithms, Adapters, and Data Distribution program (ITK-A2D2-2010), we are revisiting successful programs run by NLM in 2002 and 2004 to test the ITK-v4 major software release, to expand our research community throughout life science research in microscopy and radiology, and collect data as a foundation for imaging research in an open-science world.

We have secured the notable software development services 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. This work is funded through the American Reinvestment and Recovery Act.

Image and Text Indexing for Clinical Decision Support and Education

As part of the Clinical Information Systems effort, the Image and Text Indexing project seeks to automatically identify relevant illustrations in biomedical articles that could provide multimedia assistance to clinical decision making. We developed an experimental search engine, the Image Text Search Engine (ITSE), that augments retrieved bibliographic citations with illustrations (medical images, photographs, charts and other figures), extracted from scientific articles. The retrieval methods implemented in the search engine achieved top performance in the ImageCLEF 2010 image modality classification task.

ITSE retrieves and displays "enriched citations" - structured MEDLINE citations expanded with image-related text and concepts and linked to images and image representations based on image features. In FY2010, processes leading to generation of enriched citations were organized into a pipeline. Experiments conducted demonstrated that enriched citations improve retrieval of similar patient cases, compared to traditional citations.

Research was also conducted in key areas: ways to organize and display retrieval results using annotated images, improved methods to automatically extract both single- and multi-paneled illustrations from articles, and improved methods to extract pointers (arrows, arrowheads, symbols) within images to identify regions of interest, among others.

The project was presented to the Board of Scientific Counselors in September 2010. The Board commended the effort and also recommended providing the existing image and related text processing methods as services, and scaling the experimental search engine to large scale collections, such as PubMedCentral. To address these recommendations, researchers have started redesigning the system architecture.

Turning The Pages (TTP)

The goal of the 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, as well as 'click and turn' in an online version. The online version of TTP is a popular Web site, attracting 1,300,000 page views a month.

Following the release of the kiosk version of the Edwin Smith Papyrus, having first created a 3D model for flat scrolls, in FY2010 we released the online version that may be 'clicked and rolled out'. Commentaries on medical cases and treatments from this 1700 BC work may be viewed in synchrony with their locations in the scroll. Also this year we created the kiosk and Web versions of Hanaoka's Japanese medical manuscript. Hanaoka Seishu's fame rests on his invention of an oral anesthetic that rendered a patient unconscious for long enough to allow him to remove deep tumors.

Ongoing work includes the creation of an iPad version of TTP, with the initial release to include the Japanese manuscript, as well as Brunschwig's Liber de Arte Distillandi. Also in development are the Web and kiosk versions of *Ketab Ajaeb al-makluqat wa Gara eb al-Mawjudat (Marvelous Creatures and Mysterious Species)* compiled by al- Qazwini in the middle 1200s in what is now Iran or Iraq.

Biomedical Image Transmission via Advanced Networks (BITA)

Researchers evaluated performance of NLM-supported networks such as the Interactive Video Outreach and Distance Learning Network for Minority High School Students at Health Science-focused magnet schools, the BHEPP network linking NLM and area hospitals, 802.11a, 802.11b and higher designation wireless network implementations, and networks exhibiting narrow bandwidths, high latency and high jitter. As NLM develops applications for handheld platforms the performance of wireless networks becomes more important. In conjunction with NIH networking staff, we conducted a review of wireless infrastructure at the LHNCBC to address capacity questions raised by the movement of images within and on branch infrastructure by in-house developers of handheld applications (e.g., Turning The Pages on the iPad).

Research staff provided guidance and contacts to the organizers of an NSF-sponsored US-India Networking and e-Science Workshop held in December 2010 in India. The workshop focused on improving collaborative biomedical and science research infrastructure between the two countries using high-performance networks.

We regularly monitor the usable capacity of the existing NLM high-performance connections to the domestic and international research communities via Internet2. General capacity planning discussions internal and external to NLM have taken place focused on preparing for the movement of increasing numbers of sequencing datasets to and from NLM. staff members continue to investigate methods to measure network performance to and from NLM collaboration sites focusing on the utility of perfSONAR and other tools.

As part of this effort, research staff represented NLM and NIH on an Internet2 Advisory Council, Joint Engineering Team, Large Scale Networking Team, and other forums for high performance/speed networks.

Natural Language Processing and Text Mining

Medical Article Records System (MARS)

Automation is required to accommodate the rapid growth of the MEDLINE database, now exceeding 18,000,000 records. The MARS project aims to develop automated systems to extract bibliographic text from journal articles, in both paper as well as electronic forms. For the approximately 1,050 MEDLINE journal titles that arrive at NLM in paper form, a production MARS system combines document scanning, optical character recognition (OCR), and rule-based and machine learning algorithms to yield citation data that NLM's indexers use 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.), pattern matching to identify these entities, lexicon-based pattern matching to correct OCR errors and reduce words that are incorrectly labeled as errors to increase operator productivity.

In FY2010, LHNCBC staff provided all engineering support for the offsite MARS production facility: installation of upgraded modules, testing, maintenance and operation of all hardware and software for servers, clients and networks, and the necessary system administration. A new capability was introduced this year to accommodate up to three scanned pages in an article because we found that some lengthy abstracts require optical scanning of a third page. This required modifications to several modules, in particular those used by scanning and reconcile (verification) operators.

In addition, to help achieve the goals of NLM's Indexing 2015 Initiative, LHNCBC staff developed the Publisher Data Review (PDR) which Library Operations staff evaluated prior to deployment. The PDR system is designed to provide data missing from the XML citations sent in directly by publishers: such as databank accession numbers, NIH grant numbers, and grant support categories. By providing these missing data, PDR will reduce the manual effort in completing the citations sent in by publishers. PDR also corrects incorrect data sent in by publishers. The automated steps to fill in missing data and to correct wrong data will substantially reduce the load on the operators, eliminating the need to look through an entire article to find this information, and then to key them in. Currently, we are designing machine-learning methods to extract two additional MEDLINE citation fields: Investigator Names and Commented-on Article. If done manually, extracting names of investigators is a labor-intensive effort since articles frequently contain hundreds of such names. Similarly, identifying commented-on articles is a time-consuming process since it requires operators to read other related articles for commented-on information.

We are also developing an automated system to help operators handle cases when publishers do not supply an issue or partially supply certain citation fields to MEDLINE. This system, called WebMARS, is a software tool that operators can use to automatically create missing citations from these problematic issues. Currently, this task requires operators to manually type, copy, and paste data from online articles, a very time-consuming step.

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

Automatically Creating OldMedline Records for NLM (ACORN)

The ACORN initiative aims to capture bibliographic records from pre-1960 printed indexes (e.g., IM, QCIM, QCICL, etc.) for inclusion in the NLM OldMedline database, thereby creating a complete record of citations to the biomedical literature since Index Medicus appeared in the late 19th century. In FY2010, we continued our investigation of scanning, image enhancement, OCR, image analysis, pattern matching, and related techniques to extract unique records from the printed indexes. Finding that many of the printed indexes are available as microfilm, we decided to scan this medium rather than the paper indexes to take advantage of the lower cost of microfilm scanning. In addition, we investigated Web-based information and existing MEDLINE and OldMedline databases to avoid creating duplicate records and to correct OCR errors in citation information. Researchers designed a prototype consisting of three main components: Quality Control, Processing, and Reconcile. The Quality Control module is completed, and work is proceeding toward implementation of the Processing module which will group OCRed text into records for operators to verify using the Reconcile component. We expect to deliver the pilot system to Library Operations in FY2011.

Indexing Initiative (II)

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

The MTI system is in regular, increasing use by NLM indexers to index MEDLINE. MTI recommendations are available to them as an additional resource through the Data Creation and Maintenance System (DCMS). 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. In addition, indexing terms automatically produced by a stricter version of MTI are being used as keywords to enhance retrieval of meetings abstracts via the NLM Gateway. These meetings abstracts span the areas of AIDS/HIV, health sciences research, and space life sciences.

Indexing Initiative development focuses on improving MTI's accuracy and efficiency as well as adding functionality to the system. MTI's accuracy is increasing both incrementally based on indexer suggestions and more broadly due to system wide efforts such as incorporation of enhancements to MetaMap including negation detection and resolution of ambiguity. Recent technical enhancements have increased MetaMap's throughput three- to five-fold. One recent functional addition to MTI is an explanation facility to inform indexers how MTI arrived at specific MeSH recommendations. Another recent development consists of successfully modifying MTI both for use in NLM Cataloging and for indexing the History of Medicine's book collection. Finally, we have shown through recent experiments that MTI recommendations can be used as the initial indexing, subject to the normal review process, for a limited number of journals on which it performs exceptionally well.

Digital Preservation Research (DPR)

This project addresses an important problem for libraries and archives, viz., to retain electronic files for posterity, both documents in multiple formats (e.g., TIFF, PDF, HTML) as well as video and audio resources. Researchers focus on the preservation of digitized documents, and have built and deployed a System for Preservation of Electronic Resources (SPER). SPER builds on open source systems and standards (e.g., DSpace) while incorporating inhouse-developed modules that implement key preservation functions: ingesting, automated metadata extraction and file migration.

NLM curators are using the SPER system to preserve more than 60,000 court documents from a historic medico-legal collection acquired from the FDA. In FY2010, more than 17,000 documents were processed and added to a publicly accessible NLM Web site.

We are also investigating the applicability of SPER to preserve other historical collections of importance such as the NIAID collection of conference proceedings of the "US-Japan Cooperative Medical Science Program on Cholera Research" from 1960 to 2010. Our activities toward this initiative include: (a) recognizing different types of information within a document set through layout analysis, (b) evaluating the effectiveness of models such as Support Vector Machine and Hidden Markov Model for different metadata layouts, and (c) capturing relationships between various entities in the collection from the extracted metadata. SPER is used to perform automated metadata extraction (AME) from this document collection, and is building a related portal of research articles, authors, participating members and institutions. More robust

algorithms are being developed to enhance the AME accuracy.

In addition, research has been conducted in recognizing metadata in "Form based" documents, and was applied to prototype the extraction of metadata from the NIH Annual Reports on intramural research projects.

InfoBot

As part of the Clinical Information Systems effort, the InfoBot project enables a clinical institution to automatically augment a patient's electronic medical record (EMR) with pertinent information from NLM and other resources. InfoBot processes free-text clinical notes, extracting problems and interventions, and combining extracted information with other patient-specific information in search queries. Information sources are defined for each specific clinical task in a set of rules. Textual search results are post-processed to extract bottom-line advice and present an overview of the results at-a-glance. The manual option is useful when the API cannot be used by the EMR. An API developed for InfoBot allows integrating patient-specific information (medications linked to formularies and images of pills, evidence-based search results for patient's complaints and symptoms, MedlinePlus information for patient education) in 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 through the Evidence-Based Practice tab in CRIS since July 2009. In FY2010, we completed the evaluation of this prototype at the Clinical Center using a focus group of floor representatives of interdisciplinary teams, mainly from the nursing staff. In line with the evaluation, we modified the information delivered by InfoBot. For example, evidence search results (for a given patient) were automatically re-ranked according to the votes of other team members. We also added images of pills and a search box that allows a user to modify an automatically constructed search. The most followed links are to information about medications and protocols of clinical trials. These links are followed twice as often as the links to MedlinePlus and MEDLINE publications.

De-identification Tools

De-identification enables research on clinical documents. LHNCBC developers are designing software to de-identify clinical documents that comply with the Privacy Rule of the Health Insurance and Accountability Act of 1996. The provisions of the rule dictate removal of 18 individually identifiable health information elements that could be used to identify the individual, the individual's relatives, employers, or household members.

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

LHNCBC obtained narrative reports for approximately 70,000 individuals (under Institutional Review Board (IRB) exemptions) to develop and test this de-identification system. The record mix includes radiology reports, history and physical exam records, occupational therapy notes, discharge summaries, referrals, consult notes, laboratory data and nursing notes. These records are annotated with the assistance of a software tool that helps the annotators tag all HIPAA identifiers in the reports. In FY2010, we annotated 6,000 documents and reached a set size of 16,000 fully annotated documents for testing and improving de-identification.

The de-identification software system operates through a sequence of pipelined processes: 1) Health Level 7 (HL7) message parsing; 2) part of speech tagging; 3) protected personal health information identification, and 4) redaction. Efforts are directed toward conserving all clinically important information while ensuring none of the individually identifiable health information is included in the results set.

In FY2010, we focused our efforts on improving the system's performance on personal names. In the latest performance test of our system, we used 1,001 clinical records containing 360,879 words, of which 7,294 were personal name words. The system's overall sensitivity for correctly identifying personal name words was 99.9 percent. Researchers plan to further improve performance by incorporating HL7 header information.

Terminology Research and Services

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

Objectives for FY2011 are:

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

Medical Ontology Research (MOR)

The Medical Ontology Research (MOR) project focuses on basic research on biomedical terminologies and ontologies and their applications to natural language processing,

clinical decision support, translational medicine, data integration and interoperability.

During FY2010, staff investigated issues including quality assurance in ontologies, the representation of pharmacologic classes in biomedical terminologies, the use of the UMLS for mapping between lay terminologies and professional vocabularies, and normalization techniques for mapping clinical drug names to standard terminologies. Many of these studies leveraged the Semantic Web technologies including RDF - the Resource Description Framework - and triple stores (e.g, Virtuoso), which proved to be useful resources for integrating of biomedical information.

Researchers contributed to the LHNCBC training program by providing mentorship to one undergraduate, five graduate and one post-doctoral students, working with them on issues including data integration for pharmacogenomics studies and the representation of rare diseases in clinical and other terminologies.

In FY2010, our research activities resulted in two journal articles, 15 papers in conference proceedings, and six invited presentations. This research project was favorably reviewed by the Board of the Scientific Counselors of the Lister Hill Center. We continue to collaborate with leading ontology and terminology centers, including the National Center for Biomedical Ontology, the International Health Terminology Standards Development Organization (SNOMED CT) and the World Health Organization (ICD 11).

Clinical and Translation Science

LHNCBC is undertaking a project to facilitate the discovery of published translational science research cited in MEDLINE/PubMed, which has more than 20,000,000 citations. Using a newly-developed database of translational terms, RxNorm and MeSH, the LHNCBC developed a tool for searching for innovative, novel and promising translational research. The query is initiated using disease processes and/or interventions as search terms. Publications identified as translational in nature are then retrieved with relevant terms highlighted for easy recognition. With interventions, such as drugs in the RxNorm database and disease processes in MeSH that are pertinent clearly identified, the user can quickly find publications to facilitate research, experiment planning and bench-to-bedside applications. An additional benefit from this project is the inclusion of a new term, "translational research," to the MeSH vocabulary.

Semantic Knowledge Representation (SKR)

The Semantic Knowledge Representation (SKR) project conducts basic research in 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 the biomedical research domain and is being extended to influenza epidemic preparedness, public health, and health effects of climate change. The SKR project maintains a database of SemRep predications, which currently holds 25,000,000 predications extracted from 7,200,000 MEDLINE citations (1999 through August, 2010).

SKR efforts support innovative information management applications in biomedicine, as well as basic research. The researchers are using semantic predications to find publications that support critical questions used during the creation of clinical practice guidelines (with support from NHLBI). Semantic processing technology is being adapted to identify selected concepts in clinical narrative. One research project combines semantic predications and the literature-based discovery paradigm for investigating the physiology and pathology of sleep, and 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 and to investigate advanced statistical methods for enhanced information retrieval.

Information Resource Delivery for 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.

Clinical Research Information Systems

ClinicalTrials.gov provides the public with comprehensive information about interventional and observational clinical research studies. ClinicalTrials.gov receives over 100,000,000 page views per month and hosts approximately 900,000 unique visitors per month. At the end of FY2010, the site had nearly 97,000 protocol records, over 2,400 of which display summary results, conducted in all 50 states and in 174 countries. Approximately one-third of the studies are open to recruitment, and the remaining two-thirds are closed to recruitment or completed. Data are submitted by over 9,500 study sponsors which include the U.S. Federal government, pharmaceutical and device industries, academic, and international organizations, through a Web-based Protocol Registration System, which allows sponsors to maintain and validate information about their studies.

ClinicalTrials.gov was established by the NLM in FY2000 in response to the Food and Drug Administration Modernization Act of 1997 and to support NLM's mission of disseminating biomedical knowledge and advancing public health. Over time, ClinicalTrials.gov has been enhanced to support other registration policies and to implement the requirements of Section 801 of the Food and Drug Administration Amendments Act of 2007 [Public Law 110-85]. In FY2010, the second year after

passage of this law, new registrations were submitted at an average rate of 350 records per week. In September 2008, ClinicalTrials.gov launched the "basic results" database, which complements the registry. Registered trials may now include tables of summary results data on primary and secondary outcomes and adverse events, as well as information on the patient populations studied. Since the beginning of its operation, over 3,600 results records have been submitted by 724 study sponsors. The average number of submissions per week has increased, with an average of 50 new results records submitted per week at the end of FY2010. The expanded registration requirements as well as the results database will be further elucidated through rulemaking and NLM is working with the Food and Drug Administration (FDA) on the Notice of Proposed Rulemaking. The registry and results database provide access to information about ongoing and completed clinical research. This information is critical for clinical and policy decision makers.

In FY2010, the ClinicalTrials.gov staff was actively involved in educating the public on the new law, system requirements, and continuing to promote standards of transparency in clinical research through trial registration and results reporting. This information was communicated to a broad range of U.S. and international stakeholders via presentations and peer-reviewed publications. ClinicalTrials.gov continues to collaborate with other registries, professional organizations, and regulators in working towards developing global standards of trial registration and reporting to results databases.

Genetics Home Reference (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 almost 600 genetic conditions, more than 800 genes, 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.

Genetics Home Reference celebrated its seventh anniversary in 2010. In the past year, the project expanded its genetics content for consumers, adding more than 250 new summaries to the Web site, an increase of about 27 percent from the previous 12 months. The team also plans to continue expanding the gene families feature, which currently includes explanations of about 55 families of related genes. Usage of the GHR Web site continued to increase in FY2010. This year, the site averaged almost 16,000 visitors per day (an increase of about 13 percent over the previous fiscal year) and more than 27,600,000 hits per month (an increase of 47 percent over the previous fiscal year). GHR continues to be recognized as an important health resource.

This year, GHR staff performed outreach activities to increase public awareness of the Web site. The project continues to support the Information Rx initiative, a free program that enables doctors and nurses to write "prescriptions" directing patients to the GHR Web site for an explanation of genetic disorders and related topics. In other outreach activities, GHR staff presented the Web site to several visiting groups, including students and journalists, and represented the project at several major genetics conferences.

Profiles in Science Digital Library

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

The content of Profiles in Science is created in collaboration with the History of Medicine Division of NLM, which processes and stores the physical collections. Several collections have been donated to NLM and contain published and unpublished materials, including manuscripts, diaries, laboratory notebooks, correspondence, photographs, poems, drawings and audiovisual resources. This year, the collection of Nobel prize-winning biochemist Daniel Nathans was added to Profiles in Science. One thousand seven hundred thirty-eight transcripts of documents were also added, making handwritten items searchable and providing alternatives to PDF format files. Fifty-five digital items were also added to the thirty-three existing Profiles in Science collections. Currently 26,555 digital items composed of 139,945 image pages are available on Profiles in Science. Presently the

Web site features the archives of thirty prominent individuals.

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 FY2010, LHNCBC staff increased the reliability, security and longevity of the Profiles in Science systems. We installed a server at the NIH Consolidated Co-location Site (NCCS) to ensure availability of the Web site during facilities outages, and servers were synchronized on three different networks. We adjusted Web site update procedures to accommodate multiple servers. Staff also investigated various software for possible future use such as: JHOVE2 for validating the project's digitized files, W3C Markup Validation Service for validating Web pages, Total Validator for validating Web pages as well as Section 508 compliance, Apache PDFBox for examining PDF format files, Sitemap for easier searching by external Web crawlers, and Solr/Lucene for searching. Project staff investigated alternative formats to PDF, and found that DjVu is still the most viable alternative. Staff created a presentation, "Profiles in Science: A Digital Library," that was delivered at the 2010 International Conference on BioCommunications. We continued to develop procedures and protect the master digital files associated with the Web friendly files seen on Profiles, and to deliver the masters to collaborating institutions. We also continued to automate tasks that were being performed manually, especially those involving quality control and updating.

Evidence Based Medicine - PubMed for Handhelds

PubMed for Handhelds was developed and released in FY2003 to facilitate evidenced-based medical practice with Medline access at the point of care via smartphones, wireless PDA's, netbooks or portable laptops. PubMed for Handhelds (PMHh) 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 was recently added for a clinician’s quick reading at the point of need. New features can create a “consensus” opinion of multiple publications. Recent collaborative projects are ongoing in Botswana, Africa and the Pacific Islands. Philippines. Randomized controlled trials using simulated clinical scenarios are currently underway at the Uniformed Services University and the University of Botswana-University of Pennsylvania to evaluate the usefulness of abstracts.

Clinical Vocabulary Standards

Multiple projects in this area continue to promote the development, enhancement, and adoption of clinical vocabulary standards. The Problem List Vocabularies Project focuses on the use of controlled vocabularies in electronic problem lists. RxTerms facilitates the use of RxNorm in the capture and encoding of prescription information. Inter-terminology mapping promotes the use of standard terminologies by creating maps to administrative terminologies, thus allowing 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 CORE Problem List Subset of SNOMED CT is published in the UMLS as a specific content view. The inter-terminology maps are also available through the UMLS.

The CORE Problem List Subset of SNOMED CT

The problem list is considered to be an essential part of the Electronic Health Record (EHR) by various sanctioning bodies and medical information standards organizations, including the Institute of Medicine, Joint Commission, American Society for Testing and Materials and Health Level Seven. An encoded problem list is also one of the core objectives of the “meaningful use” regulation of EHR published by the Department of Health and Human Services. Problem lists have value beyond clinical documentation. Common uses include the generation of billing codes and clinical decision support. To drive many of these functions, an encoded problem list (as opposed to data entered as free-text) is often required. However, most institutions use their own problem list vocabularies. This lack of a common standard leads to duplication of effort and impedes data interoperability.

Based on data collected from seven large-scale U.S. and overseas healthcare institutions, a detailed study was done on the nature of the local problem list vocabularies. One significant finding is the low level of overlap between these vocabularies, with an average pairwise overlap of around 20%. However, terms that are shared among institutions were used eight times more frequently than concepts unique to one institution, which lends support to the idea of having a common core of problem list terms across institutions. Since SNOMED CT is a designated standard for problem lists according to the “meaningful use” criteria, a CORE (Clinical Observations Recording and Encoding) Problem List Subset of SNOMED CT, which contained about 6,000 concepts and represented the most frequently used problem list terms, was identified and made available to SNOMED CT users.

The CORE Subset can be used as a starter set for institutions that do not yet have a problem list vocabulary based on SNOMED CT. This will save significant development effort and reduce unintentional variations in the choice of terms. Existing problem list vocabularies can also be mapped to the CORE Subset which will facilitate data interoperability. Since publication, the CORE Subset has received considerable attention within the SNOMED CT user community. The IHTSDO (International Health Terminology Standards Development Organization) used the CORE Subset to focus its quality assurance effort on clinically important concepts. The MedlinePlus Connect Project, which facilitates online linkage to patient education information, has mapped all concepts in the CORE Subset to MedlinePlus health topics. There is ongoing effort to map the CORE Subset to the ICD classifications (ICD-10 and ICD-10-CM) which will promote the adoption of SNOMED CT by allowing re-use of SNOMED CT encoded clinical data. Currently research effort is underway to formally review the coverage and usability of the CORE Subset, in comparison to other existing problem list vocabularies.

RxTerms

Originally created for the Personal Health Record Project, RxTerms serves as an efficient drug interface terminology to facilitate electronic capture of prescription information and linkage to standard identifiers in RxNorm, the U.S. national drug terminology standard. To overcome the problem of excessively large pick-lists and long drug names in RxNorm, RxTerms segments the information contained in a RxNorm clinical drug into two portions: ingredient with route (e.g. Amoxicillin (Oral-pill)), and strength with dose form (e.g. 500 mg Tabs). This has been shown to improve data entry efficiency. RxTerms includes other usability-enhancing features like the inclusion of common synonyms and abbreviations, "tall-man lettering" to distinguish look-alike drug names and improved liquid dose concentration information. RxTerms is updated every month with the full monthly release of RxNorm.

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, researchers developed SOAP-based application programming interface (API), enabling users to integrate RxNorm in their applications. Examples of use include mapping drug names to RxNorm, finding the ingredient(s) corresponding to a brand name, and obtaining the list of NDCs for a given drug.

The RxNorm API was further developed and we designed a RESTful version of the API, compatible with the Representational State Transfer (REST) architecture. Two other drug information sources were integrated with RxNav: RxTerms, an interface terminology for prescription writing or medication history recording; and NDF-RT, a resource that links drugs to their pharmacologic classes and properties, including indications, contra-indications and drug-drug interactions.

The production version of RxNav has received about 10,000,000 queries in FY2010. Users include clinical and academic institutions, as well as pharmacy management companies, health insurance companies, EHR vendors, and drug information providers. In the future, the integration with NDF-RT will be refined. Specialized applications relying on the API will be developed, e.g., for mapping large amounts of terms and codes to RxNorm, and for crosswalk purposes between drug vocabularies through RxNorm.

Collaboration with Centers for Medicare and Medicaid Services

In collaboration with the Centers for Medicare and Medicaid Services (CMS), the LHNCBC continues to provide help and guidance regarding standard terminologies for medications and clinical problems. To assist CMS in the implementation of their Continuity Assessment Record Evaluation (CARE) data entry form, we shared with them RxTerms and the PHR Problem List Terminology. We have received some preliminary data in anonymized form captured in the CMS systems and can use that to study the adequacy of the two interface terminologies.

Standards for Identifying Clinical Observations and Orders

In FY2010, LHNCBC expanded the LOINC database. The Regenstrief Institute and the LOINC committees are committed to expanding LOINC's globalization and usability. Last spring, LOINC completed the development of a web-based LOINC browser based on the Lucene Search engine, http://search.loinc.org. LOINC currently supports nine languages including both Simplified Chinese and Korean. The most recent additions are Greek and Italian.

We are working with the major laboratory companies and the American Clinical Laboratory Association (ACLA) to clarify the content of many of the most frequently ordered test panels and mechanisms to represent them in LOINC. This has the dual advantage that the major commercial laboratories help us define the standard approach and then often change their internal systems to conform to it. Laboratory instrument vendors are now linking LOINC codes to their outputs and making the linkage available in the package insert or on their web sites. Other work that is under development is the list of the top 2,000 laboratory test result terms, with guidance about which terms to use in what situation and an improved mapping tool.

LOINC is collaborating with PhenX (https://www.phenxtoolkit.org/) and PROMIS to

incorporate their survey instruments fully within the LOINC database. Working with NHGRI, LHNCBC has enhanced the robustness and usability of the PhenX Toolkit, which allows researchers to review and select high priority measures and recommended protocols for inclusion in genome-wide association studies (GWAS) and other broad-based genomics studies. LHNCBC and NHGRI are also working together to map PhenX measures to controlled clinical vocabularies, such as LOINC and SNOMED-CT.

Newborn Screening Coding and Terminology Guide

Newborn screening is an important part of public health, but use of test results is complicated by wide variations among states in the ways tests are conducted, results recorded, and by paper-based communications. The current situation can delay rapid attention to a child's health problems, and it creates frustration and extra work for parents, health care providers, and public health authorities.

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

NLM collaborated with HRSA to create the HRSA/NLM guidance for sending electronic NBS results using HL7 messages and LOINC and SNOMED CT codes. The guidance is available on the NLM Newborn Screening Coding and Terminology Guide Web site, at http://newbornscreeningcodes.nlm.nih.gov. The goal of the guidance is to provide a standard framework for reporting newborn screening results in an electronic message whose contents can be accurately interpreted by recipient electronic health information systems for use in care, follow-up and analysis. Adoption of this framework will enable the meaningful use and comparison of data from different laboratories.

Staff worked with HRSA, CDC, the Association of Public Health Laboratories (APHL), and the National Newborn Screening and Genetics Resource Center (NNSGRC), to hold a workshop about Newborn Screening Laboratory Results and Health Information Exchange in November 2010. The goal of the workshop was to showcase implementation and adoption of the HRSA/NLM guidance for sending newborn screening results electronically using standardized universal HL7 messages and LOINC and SNOMED CT codes, distill best practices, gain feedback from the states about needed additions or edits to the guidance, and gather information for other related ongoing and planned initiatives, including developing codes for the confirmatory and diagnostic testing that occurs after newborn screening for some infants, triggered by the newborn screening results. The workshop brought together representatives from 30 states, several federal agencies, the three vendors developing and operating newborn screening laboratory information systems (PerkinElmer, Natus Medical/Neometrics and OzSystems), and members of other organizations that actively represent the newborn screening community.

Several states are using the HRSA/NLM guidance to develop standard reports of newborn screening results. We have reviewed prototype messages from all 3 of the major vendors for NBS labs that utilize the NLM/HRSA specification for reporting NBS results, and worked closely with at least 5 states (Pennsylvania, Kentucky, and New York) that are in the process of implementing the guidance. During the last year, we have been collaborating with HRSA and the Public Health Informatics Institute (PHII) about data elements and message standards related to their work on an implementation guide for sending electronic newborn screening lab orders using HL7 ver 2.x.

Communication Infrastructure Research and Tools

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

Videoconferencing and Collaboration

Researchers continued to investigate, review, and develop collaboration tools, research their application, and use the tools to support ongoing programs at the NLM. In our ongoing work with uncompressed high definition video over IP, we determined strengths and weaknesses of each of the three technologies (iHDTV, UltraGrid, and Conference XP) and continue to overcome problems encountered in the delivery of uncompressed video due to differing platforms and user machine memory capabilities.

iHDTV is the only system sufficiently robust to use in a clinical trial of uncompressed videoconferencing for telemedicine. Staff members are monitoring the progress of UltrGrid's developers and are collaborating with the developers of the uncompressed ConferenceXP program to solve the uncompressed video transmission problem. We are also monitoring the HD open source work of VLC developers that applies H.264 compression, since VLC 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. Staff have developed a manuscript comparing the main video technologies used in the Collab, including those doing standard definition video and compressed HD and some uncompressed technologies.

Staff also completed a study comparing patient, provider, and interpreter ratings of clinical encounter quality when interpretation services were provided in-person and by video and phone. In-person was rated highest and phone lowest. The study compared conventional standard definition video to phone and in person interpretation. A follow up study using lower quality video (less than full screen) and cell phone technology assessing video interpretation in pharmacy settings was started. Extensive tests were done with VSee, a low bandwidth video program.

Results of two co-location studies of videoconferencing were combined to assess learning outcomes and collaborative behaviors when students were co-located or dispersed. There were no differences in performance because all the medical students scored well on the exam, but ratings of interactivity were highest for the dispersed videoconference. This finding was attributed to the ways videoconferencing channeled communication and its realism and synchronicity.

Staff continued to work with SIS on 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 FY2010, staff conducted outreach programs with the University of North Carolina at Chapel Hill, the University of Tennessee at Memphis, the University of Maryland at Baltimore, the Virginia Commonwealth University, and the Rochester Institute of Technology.

Staff conducted a retrospective review of NLM OHPCC Telemedicine, Next Generation Internet, and Scalable Information Infrastructure Project publications and identified computing, communication, and health science application themes in this diverse corpus of work.

OHPCC Collaboratory for High Performance Computing and Communication (Collab)

The Collab was established 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 OHPCC and outside NLM. When appropriate, it is leveraged to support other activities and programs of the NLM. The facility can be configured to support a range of technologies, including 3D interactive imaging (with stereoscopic projection), the use of haptics for surgical planning and distance education, and interactive imaging and communications protocols applicable to telemedicine and distance education involving a range of interactive video and applications sharing tools. The latter 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 NLM's Visible Human and advanced networking initiatives. The collaboration technologies include a complement of tools built around the H.323 and MPEG standards for transmitting video over IP and open source technologies such as the Access Grid. Staff upgraded the H.323 technology this year and acquired 3D display and DVD playback technology that are being integrated into the remote control technology for displaying output from collaboration tools.

DocView Project: Tools for Using and Exchanging Library Information

The goal of this project is to conduct R&D on advanced tools allowing libraries and users to access biomedical information. In FY2010, research focused on the completion of MyDelivery development, including an Applications Programming Interface (API), and release of its source code. MyDelivery is a novel Internet communications system designed to deliver very large Gigabyte-sized files and large numbers of files, especially over potentially unreliable networks such as wireless used by an increasingly mobile population. It features a unique memory-based server architecture that buffers user data briefly in memory, and avoids storage of user data on server hard disk, where it could be compromised. Health science applications often require the use and exchange of information contained in very large files (e.g., digitized x-ray images, sonographic images, digital video files, MRI, CT scans, PET scans, and scanned document images). Targeted for use in clinical, research, administration, and library environments, the MyDelivery system will be capable of reliably communicating biomedical information contained in files of virtually any size over networks of all types, including potentially unreliable ones. Developers released source code and extensive system documentation for the MyDelivery client, server, and API in June 2010.

As part of the DocView project, research and system engineering continues to maintain and improve the operation of DocMorph, a Web-based server providing users remote image and information processing capabilities via the Internet. This system now accepts more than fifty file formats, including black and white images, grayscale and color images, text and word processing files, to produce four outputs: PDF files, TIFF files, text, and language translation. DocMorph averages 1,500 conversions daily, and 1,400 unique users monthly.

DocMorph's 24,000 registered users include several hundred libraries that use DocMorph as part of their interlibrary loan services. While DocMorph is generally accessed via a Web browser, the MyMorph client software allows users to perform large scale conversion of thousands of files at a time. MyMorph has more than 16,000 registered users, many of whom are document delivery librarians in small libraries around the

country, using MyMorph as an important component of their daily document delivery operation.

Computing Resources Projects

The Computing Resources (CR) Team accomplished a number of core projects to build, administer, support, and maintain an integrated and secure IT infrastructure that facilitates the research activities of the LHNCBC and thereby augments the overall effectiveness of research staff members. The integrated secure infrastructure encompasses network management, security management, facility management, storage and backup management, and system administration support for a large number of individual workstations and shared servers.

The network management team plans, implements, tests, deploys, and operates high-speed network connectivity locally as well as over Internet and Internet2. The core projects include studying and planning the implementation of central network management for effectively responding to network alerts and malfunctions; 10 Giga BPS network to support research projects that require high-speed communication capacity; and an enterprise device management system to update large number of network devices uniformly.

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

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

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

Disaster Information Management

The Lost Person Finder project, seeking to develop systems for family reunification in the aftermath of a mass casualty event, was initiated as part of the Bethesda Hospitals Emergency Preparedness Partnership (BHEPP). The systems developed in this project combine image capture, database and Web technologies, and address both hospital-based as well as community-wide scenarios.

The hospital-based LPF system includes means to photograph victims at the triage station of a hospital, and to capture these pictures and descriptive metadata (name, age range, and identifying features) in TriagePic, an application developed for the triage staff to stage the patients to appropriate treatment areas in the hospital. This data enters a MySQL database which can be searched via a Web site built by customizing the open-source Sahana disaster management system. The LPF system features a "Notification Wall" that displays images of victims on computers as well as large auditorium screens for family or staff. In 2009 and 2010, we participated in large-scale multi-institutional drills (Collaborative Multi-Agency Exercise or CMAX) and demonstrated TriagePic usage, search capability, and the Notification Wall displays at the Navy and Suburban hospitals in Bethesda.

The community-wide system was developed rapidly in January 2010 in response to the earthquake disaster in Haiti. Building on components of the existing LPF system, developers created the Haiti Earthquake People Locator (HEPL) Web site whose main element was a searchable "Interactive Notification Wall" displaying pictures and metadata of missing as well as found people. This data could be sent to HEPL via computer, cell phone, and an iPhone application, ReUnite, developed in-house and made freely available through Apple's iTunes (ReUnite, NLM's first iPhone app, may also be used with iPod Touch and iPad, and is available for any future disaster).

The engineering effort to develop HEPL involved collaborations throughout the LHNCBC and other NLM divisions. Since it was necessary for HEPL to be understood in English, French and Kreyol, language translations were drawn from various NLM divisions as well as from DC Crisis Camps. Initially, HEPL acquired missing person data by scraping CNN iReport records, and later our system was made interoperable with the Google People Finder site. Eventually, more than 50,000 records (photos and metadata) were made searchable through HEPL. HEPL also offered links to the Google site, metasearch engines, and Haiti-specific content created by NLM's Specialized Information Systems division.

In an effort to provide a system capable of being used to respond to any disaster anywhere, developers are designing a prototype unified system, NLM Person Locator (PL), to hold data from multiple disasters, thereby eliminating the need to build multiple Web site/database instances.

Since personally identifiable information (PII) such as pictures and names are collected by LPF, staff have initiated a Certification and Accreditation Process and developed a System Security Plan. An extension of the original provisional OPM data collection permission has been granted approval through 2013.

Video Production, Retrieval, and Reuse Project

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

The still image, graphics, and video support staff provide ongoing capability to all of the NLM. This work includes production, post-production, and authoring services for the development of Internet video, kiosk interactive multimedia, and DVDs. This area of focus includes support to maintain the audio, video, and multimedia capability in the NLM board room, auditorium, and other conference areas.

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

The 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. Extensive development work continued toward the planning and demonstration of interactive multimedia for the FY2011 NLM Exhibition "Native Concepts of Health and Illness." As an extension of this work, staff developed prototype iPhone and iPad applications. Based on these prototypes, the NLM plans to integrate iPhone applications into the upcoming exhibition.

Training Opportunities

Working towards the future of biomedical informatics research and development, the LHNCBC provides training and mentorship for individuals at various stages in their careers. The LHNCBC Informatics Training Program (ITP), ranging from a few months to more than a year, is available for visiting scientists and students. Each fellow is matched with a mentor from the research staff and participates actively on Center research projects.

In FY2010, the LHNCBC provided training to 51 participants from 15 states and seven countries. Participants worked on research projects including 3-D informatics, clinical information systems, content-based information retrieval, de-identification of medical records; image, text and document processing, information retrieval research, interactive publication research, medical ontology research, medical terminology research, mobile computing, natural language processing, personal health record and telemedicine projects.

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

The ITP sponsors a Clinical Informatics Postdoctoral Fellowship Program to attract young physicians to NIH to pursue research in informatics. This program is run jointly by the LHNCBC and the Clinical Center to bring postdoctoral fellows to labs throughout NIH, though funding is from the LHNCBC. We continue 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 the eight-week NLM Rotation Program which provides trainees from NLM-funded Medical Informatics programs with an opportunity to learn about NLM programs and current LHNCBC research. The rotation includes a series of lectures covering research being conducted at NLM and the 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 in November 1988 by Public Law 100-607 as a division of the National Library of Medicine. The establishment of the NCBI by Congress reflected the important role information science and computer technology play in helping to elucidate and understand the molecular processes that control health and disease. Since the Center's inception in 1988, NCBI has established itself as a leading resource, both nationally and internationally, for molecular biology information.

NCBI is charged with providing access to public data and analysis tools for studying molecular biology information. Over the past 22 years, the ability to integrate vast amounts of complex and diverse biological information created the scientific discipline of bioinformatics. The flood of genomic data, most notably gene sequence and mapping information, has played a large role in the increased use of bioinformatics. Recently, next generation sequencing has been a source of large amounts of 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 supports a multidisciplinary staff of senior scientists, postdoctoral fellows, and support personnel. NCBI scientists have backgrounds in medicine, molecular biology, biochemistry, genetics, biophysics, structural biology, computer and information science, and mathematics. These multidisciplinary researchers conduct studies in computational biology and apply the results of their research to the development of public information resources.

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

In order to fulfill its mission, NCBI:

- Creates automated systems for storing and analyzing information about molecular biology and genetics combined with evidence in biomedical literature;
- Performs research into advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules and compounds;
- Facilitates the use of databases and software by researchers and healthcare personnel; and
- 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 research in genomic analysis.

GenBank

The basis for NCBI sequence data is GenBank, the NIH genetic sequence database. GenBank is an annotated collection of all publicly available DNA sequences. NCBI is responsible for all phases of GenBank production, support, and distribution, including timely and accurate processing of sequence records and biological review of both new sequence entries and updates to existing entries.

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

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

The traditional nucleotide database is divided as well into three specialized components consisting of Expressed Sequence Tags (ESTs), Genome Survey Sequence (GSS) records, and the "CoreNucleotide" group. The Transcriptome Shotgun Assembly (TSA)

division contains shotgun assemblies of primary (mRNA) sequences deposited in dbEST, the Trace Archive, or the Short-Read Archive (SRA).

The Third Party Annotation (TPA) database supports third party annotation of sequence data already available in public databases. In order to be included in the TPA database, the analyses must be published in a peer-reviewed scientific journal. TPA records are divided into two sections, TPA:experimental and TPA:inferential. TPA:experimental contains data supported by peer-reviewed, experimental evidence. TPA:inferential contains data by inference where the source molecule or its products have not been the direct result of experimentation, and are often computationally derived. New full releases of GenBank are distributed every two months. Daily updates are made available via the Internet and the World Wide Web.

The amount of data submitted to GenBank grows continuously. In fact, from 1982 to the present, the number of bases in GenBank has doubled approximately every 18 months. GenBank's two divisions combined contain over 181 million sequence records and over 286 billion base pairs. The traditional nucleotide sequences division increased to 122 million records in FY2010 from 108 million in FY2009. The WGS division has grown to over 59 million records and 169 billion base pairs. About 200 complete microbial genomes were processed by NCBI this 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.

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 NCBI and outside resources, such as biological databases and sequencing centers. The links enable GenBank to serve as a key component in an integrated database system that allows researchers to perform comprehensive and seamless searching across all related biological data on the NCBI website.

NCBI has developed various tools for GenBank data submission. Sequin is a stand-alone tool that updates and submits large groups of sequences. This year, a new version of NCBI's Web-based submission tool, BankIt, was released with new modifications. New BankIt allows submitters to upload and annotate multiple sequences at once and resume a previous submission at a later time. These are marked improvements over the previous version of BankIt, which required submitters to annotate and submit their sequence records one at a time and in a single session. In addition, new tabs are available at the top of submission pages to navigate and edit previously visited pages. Guides for specialized submissions such as genomes, batch sequences, and alignments are available online.

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 dbSNP, RefSeq, CCDS, dbGaP, Entrez Gene, Probe, UniGene, HomoloGene, dbVar, 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 curated RefSeq collection contains 11,223,870 proteins representing 10,854 organisms. This represents a 27 percent increase in the number of proteins from last year. The RefSeq curation group supports the whole genome annotation process flow for updating existing genomes and processing new submissions.

The RefSeqGene project, a subset of the RefSeq project, provides a set of genomic sequences of well-characterized genes to use as a reference standard. Over the past year, the number of RefSeqGene records grew to 3,418. Activity within the Locus Reference Genomic collaboration increased with 200 RefSeqGenes assigned LRG accessions, with 54 of these made 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 in a timely fashion. Descriptions of variants were extracted from OMIM, GeneReviews, and GeneTests, and processed for submission to dbSNP. The ongoing project will have all variants represented by OMIM assigned identifiers in dbSNP and dbVar by the end of FY2010.

The Consensus CoDing Sequence (CCDS) database identifies a core set of consistently annotated, high-quality human and mouse protein coding regions. This year, over 800 entries were subject to collaborative

review. Updates are reviewed and must be agreed upon by members of an international collaboration. Public annotation is provided to explain the evidence that supported the change.

The Entrez Gene database provides a unified query environment for genes defined by sequence and/or genes included in the Map Viewer. It integrates information about genes and gene features annotated in RefSeq and collaborating model organism databases. The database continues to be heavily used and manages information for more than 6.5 million genes from over 6,660 taxa. During FY2010, most development was focused on internal infrastructure, namely restructuring the public display of the Gene database to make it consistent with PubMed and sequence databases. The improvements enhance user experience as well as reduce maintenance costs due to many components being provided for multiple NCBI resources.

Extensive testing, quality assurance, and documentation are essential to the release of data in Entrez Gene, Map Viewer, and BLAST, as well as documentation for Web sites that support the scientific community's access and use of NCBI resources. In FY2010, several new genomes were annotated and updates were provided to existing genome assemblies for several eukaryotic species including human, *Saccoglossus kowalevskii* (acorn worm), *Danio rerio* (zebrafish), *Xenopus tropicalis* (frog), *Oryctolagus cuniculus* (rabbit), *Bos taurus* (cow), *Rattus norvegicus* (rat), *Callithrix jacchus* (marmoset), *Macaca mulatta* (rhesus macaque), *Pongo abelii* (orangutan), *Ailuropoda melanoleuca* (giant panda).

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. 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.

The Evidence Viewer is a Map Viewer feature that provides graphical biological evidence supporting a particular gene model. The Model Maker tool allows users to build a gene model using selected exons. The NCBI Remap tool, released in FY2010, allows users to remap genome coordinates from one assembly to another. The remapping uses the alignments of one assembly to the other to project annotation features.

The Genome Reference Consortium (GRC) is an international collaboration that aims to update and improve the mouse and human genome assemblies. NCBI provides informatics support for the project such as tracking of tiling path files, overlaps between adjacent clones, and curation, and generates the final assembly after collaboration and quality assurance. In FY2010, GRC released a new version of the human genome (GRCh37) and two minor patch updates.

The UniGene database provides non-redundant clusters for the highly redundant sets of transcript sequences of expressed genes. Interactive usage of UniGene remains steady with 12,000 Web hits daily, 5,000 of which involve reviewing clusters. While UniGene has expanded to include 128 organisms, two-thirds of interactive traffic is for human and mouse entries. UniGene continues to explore opportunities to incorporate 454 and other types of short reads into its processing, most likely in the form of assemblies of transcribed sequences.

The database of single nucleotide polymorphisms (dbSNP) is a comprehensive catalog of common genetic variation. dbSNP contains over 143 million submissions of human genome data that has been processed and reduced to a non-redundant set of 30 million refSNP clusters. Eighty-four other organisms (up from 54 one year ago) are represented in the SNP database, with 98 million submissions curated to 46 million refSNP clusters.

The Probe database, part of the Entrez system, stores molecular probe data, together with information on success or failure of the probes in different experimental contexts. Nucleic acid probes are molecules that complement a specific gene transcript or DNA sequence and are useful in gene silencing, genome mapping, and genome variation analysis. The database contains over 10 million probes as of October 2010. The RNA interference (RNAi) resource stores the sequences of RNAi reagents and experimental results using those reagents, such as the extent of gene silencing and a variety of phenotypic observations.

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

The UniVec database is used to quickly indentify segments of nucleic acid sequences that are of vector origin or vector contamination. Version 5.2 of the database included a 3 percent increase in the number of sequences. The vector BLAST database was updated to contain full-length versions of all sequences in GenBank used to build Version 5.2.

Comparative Genome Data

NCBI provides guides for comparing organisms on a genome scale. There are currently 37 guides available. The Genome Resource Guides provide information on genome-related tools and repositories available through NCBI and various outside centers and institutions. The guides provide easy navigation to NCBI resources, such as organism-specific BLAST and Map Viewer pages, and list outside resources that provide sequence, mapping, and clone information. The Guides also list documentation, annotation, and comparative genomic projects. New

genomes were annotated and annotation updates were provided to existing genome assemblies for several eukaryotic vertebrate species including *Saccoglossus kowalevskii* (acorn worm), *Danio rerio* (zebrafish), *Xenopus tropicalis* (frog), *Oryctolagus cuniculus* (rabbit), *Bos taurus* (cow), *Rattus norvegicus* (rat), *Callithrix jacchus* (marmoset), *Macaca mulatta* (rhesus macaque), *Pongo abelii* (orangutan), *Ailuropoda melanoleuca* (giant panda).

The Entrez Genome database provides views for a variety of genomes, complete chromosomes, sequence maps with contigs, and integrated genetic and physical maps. The database is organized into six major organism groups: Archaea, Bacteria, Eukaryotae, Viruses, Viroids, and Plasmids and includes complete chromosomes, organelles and plasmids, as well as draft genome assemblies. It includes, but is not limited to, assembly, annotation, and genome sequencing projects, such as whole genome shotgun or BAC ends, large-scale EST, and cDNA projects. The six organism-specific overviews function as portals from which all projects in the database pertaining to that organism can be browsed and retrieved. There are currently 1,046 eukaryotic genome sequencing projects completed, in draft form or in progress.

The Viral Genomes Web site provides a convenient way to retrieve, view, and analyze complete genomes of viruses and phages. NCBI's viral genotyping tool helps identify the genotype of a viral sequence using BLAST. NCBI currently provides access to 3,738 reference sequences for 2,554 viral genomes and 41 reference sequences for viroids.

Fungal Genomes Central is a portal to information and resources about fungi and fungal sequencing projects. There are currently 125 fungal genomes in various stages of annotation. Plant Genomes Central is an integrated, Web-based portal to plant genomics data and tools. It provides access to large-scale genomic and EST sequencing projects and high resolution mapping projects. Fifty-nine plant species are represented in the genome project database.

The Microbial Genome Annotation Pipeline was developed for annotation of prokaryotic genomes. Over 1,428 genomes have been annotated in-house and NCBI is working with ten outside groups who submit data.

Specialized Databases and Tools

NCBI has begun development of databases to represent a more comprehensive registry of genetic tests, entitled the Genetic Testing Registry, and medically important sequence variation, entitled ClinVar. A beta version of ClinVar was available for the meeting of the American Society of Human Genetics in November 2010. The Genetic Testing Registry (GTR) will provide a centralized online resource for information about the availability and scientific basis of genetic tests. The GTR is scheduled to launch in 2011.

The database of genomic structural variation (dbVar) contains data on variant DNA less than or equal to one basepair in size. Submissions are accepted from whole genome comparative studies and locus- and gene-specific data from quantitative studies. In FY2010, dbVar began accepting submissions and made available a site to facilitate submissions, obtain documentation, and provide FTP access to loaded studies. The database was formally released and integrated into the Entrez system late this year. Over 50 studies from seven organisms have been loaded into the database.

The Influenza Virus Resource is a comprehensive collection of flu sequences. Samples collected all over the world include viruses obtained from birds, pigs, humans, and other species. Data is obtained from the NIAID Influenza Genome Sequencing Project, GenBank, and more than 60 other institutions worldwide. Links are provided to other flu resources containing sequences, publications, and general flu virus information. Over 39,600 new influenza virus sequences were entered into NCBI's Influenza Sequence Database in FY2010.

The Flu Dataset Explorer provides an interactive tool for preliminary analysis of protein sequences from the Influenza Sequence Database or from a user's own file. New functionalities added to the Influenza Virus Resource include the ability to: search sequences by drug-resistant mutations, GenBank release date and multiple GenBank accession numbers; customize definition lines of FASTA sequences; and retrieve sequences only when selected other segments of the same virus also exist in the database.

The NCBI Trace Archive is a permanent repository of DNA sequence chromatograms (traces), base calls, and quality estimates for single-pass reads from various large-scale sequencing projects. The trace data can be scanned using a rapid nucleotide-level cross-species sequence similarity search program called cross-species MegaBLAST. Using the visualization tools of the related Assembly Archive, researchers can examine an assembly of trace data from which a finished genomic nucleotide sequence has been derived. They can determine, for instance, if a crucial nucleotide base change associated with a disease is supported by the sequence evidence. The Trace Archive currently holds over 2.1 billion traces representing over 980 species.

The Sequence Read Archive (SRA) archives data generated from massively parallel sequencing experiments. SRA presents data by sequencer runs rather than individual traces and currently contains 4,771 studies and 100 terabytes of data. SRA WGS sequences are now searchable through the SRA BLAST page.

The Clone Registry is a database that integrates information about genomic clones and libraries, including sequence data, genomic position, and distributor information. Management of information about clones of interest to users of Gene and UniGene is being consolidated in the Clone Registry.

The Gene Expression Omnibus, or GEO, is a high-throughput gene expression/molecular abundance data repository providing curated online storage and retrieval of gene expression data. Profiles are submitted via GEOarchive, a spreadsheet format for large batch submissions. GEO received over 145,000 new submissions in FY2010, and now has more than 580,000 records. Ten billion new individual data points were also accepted this year, bringing the total number to over 33 billion. More than 3,000 new manuscripts cited GEO accessions or provided a direct reference to the database, for a current total of over 10,000 manuscripts. And, an additional 500 new curated GEO Dataset records were created and released, bringing the total to 2,722. As with other NCBI databases, GEO was modified to better represent next-generation sequence submissions.

Data from the NIH Roadmap Epigenomics Mapping Consortium are being deposited into GEO. The NCBI Epigenomics database serves as a comprehensive resource for whole-genome epigenetic datasets. Epigenetics is the study of stable and heritable changes in gene expression that occur independently of the primary DNA sequence. The Epigenomics database has been constructed by selecting the subset of epigenetics-specific data from general-purpose archives such as GEO and SRA, then subjecting them to further review, annotation, and reorganization. The Epigenomics database contains 66 studies, 318 samples, and over 1,100 data tracks from five well-studied species. A new home page was unveiled this year following the new Entrez design.

A dedicated Roadmap Epigenomics Data Listings page was built and released in FY2010, allowing for simple data downloads as well as data visualization in NCBI's sequence viewer. A Sample Browser for viewing available samples was also released, as well as an Advanced Search tool for more extensive data mining. Documentation for using the resource, managing collections, and viewing and downloading genome tracks was added to the site.

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

The UniVec database is used to quickly identify segments of nucleic acid sequences that are of vector origin or vector contamination. Version 5.2, released in FY2010, included a 3 percent increase in the number of sequences. The related Vector BLAST database was updated to contain full-length versions of all sequences of GenBank used to build version 5.2.

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. Four releases were provided to the public in FY2010 and average monthly usage rose by almost 300 percent. In the same time period, the application was downloaded 8,900 times, including at least one full download of the source code for customized installation. Much of the development and visualization of Genome Workbench transfers directly to its Web-based counterpart, Sequence Viewer, which saw approximately 340 percent growth in usage within one year. Sequence Viewer is designed to be an embeddable component to compliment other information-rich views and there is currently one application using it as such. In FY2010, new tutorials were developed and released, giving users up-to-date examples and documentation.

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

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.

The Peptidome resource is a repository that archives and distributes tandem mass spectrometry peptide and protein identification data. Mass spectrometry results and conclusion-level information are captured, together with sufficient raw data and descriptive information to enable understanding of the experiment and analysis of the underlying data. The database contains 44 studies (up from 23) with information from 312 biological samples, a 63 percent increase from FY2009. 60,980 unique proteins and 267,773 unique peptides from over 3.9 million individual mass spectrometry spectra have been processed from 952GB of submitted data. In addition, Peptidome increased to over 52 million MS spectra (from 8.4 million one year ago) and over 48 million MS2 spectra (from 21.4 million).

Chemical Information

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

PubChem BioAssay allows users to examine descriptions of each assay's parameters and readouts, with links to substances and compounds enhanced by a queuing system and caching mechanism. The number of bioassay records markedly increased from 1,750 in September 2009 to over 462,600 in August 2010. The majority of this increase was the result of importing the literature abstracted bioactivity data from the EBI ChEMBL database. Year-to-year growth of new depositors for both databases was over 20 percent. The volume of biological data increased by 43 million test results in FY2010, up 81 percent from last year, for a total of 96 million. Forty million, or 92 percent, of this increase came from the NIH Molecular Libraries Program (MLP) screening center network.

The PubChem Compound database provides unique chemical structures and validated chemical depiction information describing substances in the PubChem Substance database. PubChem Compound grew by three million records, a 12 percent increase. The PubChem Substance database contains chemical substance records and associated information. PubChem Substance grew by 11 million records, an 18 percent increase. The Web site is now visited by more than 60,000 users per day, a 20 percent increase from last year.

As part of the Entrez system, PubChem contains an extensive set of links to related information within its own sets of data as well as to other Entrez databases and outside resources. PubChem interfaces were updated to include rapid drill down tools, subset selection, and data analysis facilities. To keep up with PubChem's continued and steady growth, major infrastructure improvements were carried out. These upgrades are intended to:

help streamline updates to Entrez, the PubChem FTP site, and production Web sites; enable larger data analyses to be performed by users; allow larger Unique Identifier (UID) lists to be managed; and allow PubChem services to operate simultaneously at the NCBI co-location facility. New tools are being developed to further improve upon structure-function analysis by using, for example, multiple 3D conformers, cross-array target-based data analysis, and developing and augmenting existing tools for RNAi screens.

Protein Structure

The NCBI BioSystems database connects biosystem records with associated literature, molecular, and chemical data and serves to facilitate computation on biosystems data. The database now mirrors metabolic pathways and similar biological systems as imported from KEGG, BioCyc (first- and second-tier pathways), Reactome, and NCI's Pathway Interaction Database (PID). BioSystems is linked to Entrez Gene, Entrez Protein, PubChem, PubMed, and NCBI Taxonomy. It provides Entrez users with a means to explore functional relationships between genes, proteins, and small molecules.

FLink, a newly developed tool that lets users of Entrez tabulate links between databases together with, and sorted by, link frequency, has been released to the public. FLink addresses the needs of researchers who collect and interpret gene expression and proteomics data, and are interested in discovering commonalities and patterns in large data sets.

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

MMDB contains over 50,000 unique, experimentally determined 3D structure records. The database is updated weekly, with the source PDB data checked for consistency in the purported chemistry, sequence, and 3D coordinates. MMDB now computes and stores molecular interactions observed in three-dimensional structure complexes. In upcoming versions of the MMDB summary pages, these interactions will be displayed prominently to emphasize information about the biological function of those complexes. The presentation of MMDB has been enhanced and now features a "Selected Structures" dialog. "Selected Structures" enables users to subset the results of their queries by criteria that use computed associations with other data.

An interaction tracking database and corresponding Web service termed IBIS (Inferred Biomolecular Interactions Server) has been updated on a regular basis as MMDB continues to grow. IBIS emphasizes and displays molecular interactions that have been observed multiple times in independent experiments, and clusters such recurring observations. Molecular interactions and the corresponding sites on proteins can also be inferred for many proteins in NCBI's database, which have significant similarity to structures with known interactions. IBIS displays and infers interactions between proteins, proteins and nucleic acids, and proteins and chemicals. Protein interactions with ions were added in FY2010.

The Conserved Domain Database (CDD) is the Entrez database of sequence alignments and profiles defining protein domains as recurrent evolutionary modules, ancient domains, and full-length proteins. The CDD annotation staff produces curated hierarchies of

models related by descent from a common ancestor, representing the ancient evolutionary history of protein and domain families.

CD-Search is the Web interface to CDD, which visualizes domain-based annotation on protein sequences. CD-Search also shows the location of conserved sites, as inferred from alignments to NCBI-curated domain models, such as active sites and binding sites, providing direct links to the domain models of the site data and evidence collected by CDD curators. A batch version of the CD-Search Web application was developed and released to beta testers. CD-Search computes and displays annotation of conserved domain footprints and associated functional sites for protein query sequences. The batch CD-Search can process up to 10,000 queries at a time, and return results in various formats.

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

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

VAST, or the Vector Alignment Search Tool, is a service that identifies similar three-dimensional structures of newly determined proteins. Preliminary work has been completed on a revised version of the VAST structure neighbor display, which will emphasize overall similarity of three-dimensional structures and structure complexes. This facilitates the representation of 3D structure relationships as intra-database links at the MMDB/Structures database level in Entrez, and will render the 3D-Domains database obsolete, simplifying navigation for users with an interest in exploring 3D structure.

BLAST Suite of Sequence Comparison Programs

Comparison, whether of morphological features or protein and DNA sequences, lies at the heart of biology. BLAST has made it easier to rapidly scan huge sequence databases for similar sequences and to statistically evaluate the resulting matches. In a matter of seconds, BLAST compares a user's sequence with millions of known sequences and determines the closest matches. The NCBI Web interface for BLAST allows users to assign titles to searches, to review recent search results, and to save parameter sets in My NCBI for future use.

The BLAST suite of programs is continuously enhanced for effectiveness and ease of use. BLAST search pages were updated with the ability to exclude model sequences and environmental samples, as well as general classes to sequences. A global alignment page was also released. Three updates to BLAST+ suite of tools, as well as older C toolkit versions, were released in FY2010. Those updated included support for translated subject masking, a new blast_formatter application, and performance improvements for client software.

The COBALT multiple-sequence alignment page was improved to allow downloading of results to a local file in various formats such as FASTA plus gaps, ClustalW, Phylip, and Nexus.

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

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

Integration of Clinical, Genetic, and Environmental Databases

NCBI's database of Genotypes and Phenotypes, dbGaP, is a permanent archive originally constructed to house, display and distribute data from Genome Wide Association Studies (GWAS) with human study participants. The data from GWAS studies can typically be categorized into two main classes of data: 1) Phenotype data, consisting of clinical, anthropomorphic, demographic and exposure variables collected from the study participants; and (2) the genotypes produced using so called GWAS chip arrays, which assay from 100,000 to 1 million or more Single Nucleotide Polymorphisms (SNPs). For many studies, there also exists important documentation on how data or samples were collected, collectively referred to as study documents, herein. In addition to archiving the data, dbGaP provides public pages for scientists to view variables summaries and study specific information such as participant inclusion /exclusion criteria, attribution, and a listing of controlled access users.

Although dbGaP was originally built for GWAS data, the system was readily adaptable to accept data for many different types of studies that include human phenotype information. In the last 18 months, studies consisting of high throughput sequencing studies have contributed an increasing fraction of the datasets registered in dbGaP.

Study Submissions

dbGaP has released 109 studies to date. The studies contain public summary-level data and/or individual level phenotype/genotype data and/or high throughput sequencing data distributed through the authorized access (AA) system. More than 50 of the 109 studies were released during the year. Notable studies released in FY2010 include five large longitudinal cohort studies from the NHLBI-funded Candidate Gene Association Resource (CARe) project, and 14 meta-genomic studies from the NIH Roadmap-funded Human Microbiome Project, as well as several Gene Environment Association case controlled studies funded by NHGRI.

Each study submitted to dbGaP has disease traits in the form of Medical Subject Headings (MeSH) terms attached by the submitting investigator. Collectively, the studies released in 2010 included measurements for 283,000 total research participants; measurements included:

- Over 100,000 individual-level traits (dbGaP variables);
- 1,270 documents linked to 102,940 variable summaries. These document the studies through natural language descriptions, collection forms, or scientific protocols, and also provide models for future research. Links between searchable documents and variable descriptions provide an unprecedented level of functionality and usability;
- Millions of additional individual genotypes, i.e., single-nucleotide measurements of participant DNA sequence. These potentially reveal systematic and heritable genetic differences between affected and unaffected individuals; and
- 2,819 pre-computed statistical associations between select phenotype traits and participant genotypes describe locations in the human genome where differences between affected and unaffected participants are statistically significant.

Phenotype variable summary metrics, data dictionaries, summary-level association results and study document XML files are distributed to the public via the dbGaP FTP service.

Authorized Access System Download Activity

The dbGaP authorized access system is the NCBI portal where Principal Investigators (PI) log in and request access to de-identified individual-level data archived in dbGaP. 1,579 PI-defined research projects have been created in the approval system as of October 2010. 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.

After being counter-signed by an institutional signing official (SO), the DARs are routed to one or more NIH Institute/Center Data Access Committee (DAC) for review. DAC review confirms that each proposed research use is consistent with the data use limitations placed on use of the data by study participants during the informed consent process. Approved users return to the dbGaP system to create and download a password-protected copy of the de-identified individual-level data (e.g. phenotypes and genotypes) to their local secure computing environment.

Data Usability: Tools and Software Development

During 2010, there was significant progress in the development of GaP Plus, an association-result browser that will allow researchers to quickly drill down to results by genomic location and/or phenotypic trait. The GaP Plus interface will incorporate information curated by the NHGRI GWAS catalog (Hindroff et. al., 2009) of published Genome Wide Association Study results.

Entrez Retrieval System

Entrez, the major database search, retrieval, and indexing system at NCBI, was originally developed for searching nucleotide and protein sequence databases and related MEDLINE citations, but has since expanded to become the indexing and search foundation for all of NCBI's major resources. With Entrez, users quickly and easily search gigabytes of sequence and literature data. A key feature of the system is the concept of "neighboring," which automatically identifies references or sequences that are related to a user's query. 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 42 databases, providing sequence, taxonomy, gene, chemical, and biomedical literature and data.

Discovery Initiative

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

The Discovery Project has introduced many new features to Entrez to enhance user discovery of information that may be relevant to them. Various ads on results pages provide links to related articles and citations, queries, gene names, gene symbols, free full text, protein and nucleotide records, and accession numbers. Providing access to varied information has enabled users to find data beyond a simple text search.

Sensors are a discovery component that detects certain search terms and provides access to relevant results. For example, the Gene Sensor shows a gene symbol when it matches the current query, and provides a link to the entry in the Gene database. A Hot Topic Sensor appears for searches relevant to current topical issues such as H1N1 viral sequences. This year, NCBI began incorporating sensors and other discovery-related features into Entrez databases beyond PubMed.

The Structure database added a new "Selected Structures" filter to search results pages. This feature provides a way of sorting results into subsets based on characteristics of retrieved records. The top five protein domain families found among the retrieved structures are listed, inferring protein function, as well as the top five organisms represented in the structures. Available subsets are keyed to protein domain families, source organisms, the presence of specific molecular complexes, and presence of links to literature and taxonomy. A structure count for each subset allows users to see a complete view of available related data.

New sequence database pages (Nucleotide, Protein, GSS, EST) were released this year incorporating many discovery components. A "Results" filter provides a glimpse of the number of records present in specific categories along with links for viewing. A "Top Organisms" feature displays the number of records available for specific taxa with an option to view a taxonomic group tree. A "Find related data" feature provides a pull-down menu of other databases in which to find related information. An important new feature is the option to download coding sequence regions of various types and lengths. This often-requested feature will allow users to easily create a local database of sequences for analysis.

The NCBI and PubMed homepages were completely redesigned in FY2010. In addition to a more streamlined look, the new NCBI homepage functions as a Guide Site that provides rapid access to all resources and tools offered by NCBI. Resources are divided into categories and further into functional sections such as Databases, Tools, Submission, and Downloads. A new "How To" section provides short explanations on using resources and finding information on the NCBI Web site.

The new PubMed interface has a more basic format with the same functionality as the old pages. Tabs were replaced with a powerful Advanced Search page. All redesigned pages contain a new search bar and footer area that will eventually be standard on all NCBI pages. The new footer provides rapid navigation to major areas of the NCBI Web site. In the coming year, more Entrez databases will adopt the new PubMed format.

Literature Information Resources

PubMed

PubMed provides Web-based access to citations and abstracts for the biomedical science journal literature. PubMed is comprised primarily of journals indexed in NLM's MEDLINE database, but also contains a limited number of journals outside the scope of MEDLINE. Links to articles available in full text through NCBI's PubMed Central database are also provided. Serving as the foundation of NCBI's bibliographic information system, PubMed contains over 20.9 million citations from more than 36,772 journals, some dating back to the 1950s.

PubMed is continually updated and enhanced for better functionality and more precise search results. The PubMed homepage was redesigned this year with a three-column design containing categories for improved ease of use. The PubMed interface changed as well: the "Send To" and Clipboard options are more visible; "Filter your results" and "Manage filters" replaced the "Filter" tabs; Limits options can be activated on the new Advanced Search page; and "Find related data" replaced the Links option. A "Books and Documents" feature was added to search results, linking to book chapters in the NCBI Bookshelf when available. "Display formats" and "Display settings" were modified and an "Auto Suggest" feature was included on the query page.

The "Advanced Search" page was modified to provide users with a better method to build searches. For example, the search box was streamlined with "Search" and "Preview" buttons to process terms present in either the "Search Builder" box or main "Search" box.

"My NCBI" is an Entrez feature that allows users to store searches and results. It also provides the option of automatically updating searches and sending results via email. PubMed results within My NCBI can be tailored to each user's preference with the new "Results Display Settings" option. My NCBI now supports the use of OpenID and InCommon Federation accounts for log in information. As of September 2010, 33 federated sites are accepted including Google, PayPal, and eRA Login as well as many universities. My Bibliography now allows users to manage compliance with the NIH Public Access Policy as well as add citations from books, meetings, presentations, patents, and articles not found in PubMed.

PubMed 2010 DTDs went into effect in December 2009. E-Utilities also enacted a slight policy change for some parameters. E-Utilities documentation was added to the Bookshelf as a chapter book.

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

During 2010, the number of organizations participating in LinkOut increased to over 3,070, representing an 8 percent growth rate over the past year. LinkOut participants include 2,403 libraries, over 390 full-text providers, and 275 providers of non-bibliographic resources, such as biological and chemical databases. Users can now link to 94 million Entrez records, including links to the full text of 56 percent of PubMed records from over 8,300 journals.

Outside Tool is a related service that also links users to outside resources. Participation in this program increased to over 730 institutions. Usage of LinkOut resources reached over 33 million hits per month, and about 1.3 million hits per week day.

LinkOut indexing is being redeveloped and in the final stage of testing. The goals of redevelopment are to: support Unicode and languages beyond English, and to adopt a modular approach to LinkOut indexing to improve its efficiency and reliability. A LinkOut portlet displays LinkOut resources more prominently in PubMed and in other Entrez databases to display LinkOut resources. A number of enhancements have been made to the library submission utility. A portal application for libraries to check their linking information independently has been developed. An automatic link-checking procedure has been instituted as a regular process in order to be proactive in improving the quality of links.

PubMed Central

PubMed Central (PMC) archives, indexes, and provides free and unrestricted access to full-text articles from life science journals. This repository is integrated with the PubMed biomedical literature database of indexed citations and abstracts. During FY2010, PMC celebrated its tenth anniversary with a one-day symposium featuring speakers from NCBI and the outside community.

PMC now has more than 900 participating journals. Between them, they deposited more than 25 percent of the estimated 88,000 NIH-funded articles published in the last year. In June 2010, the total number of articles in PMC reached the 2 million mark. More than 37,000 author manuscripts of NIH-funded articles were deposited in PMC via the NIH Manuscript Submission System. Use of PMC continues to increase in concert with the growth of available articles. On a typical weekday, PMC has up to 420,000 unique users retrieving over 700,000 articles.

A new collaborating site, PMC Canada, began operating at the end of 2009. It joined UKPMC as a member of the PMC International (PMCI) Network.

Bookshelf

The NCBI Bookshelf gives users access to the full text of over 600 textbooks 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. This year, 148 new books were added to the Bookshelf, more than any other previous year. Among the new titles are five clinical guidelines from the NIH and Clinical Excellence (UK), Stembook, and Health United States 2009. Over 200 chapters were added to existing books and documentation, including GeneReviews, MICAD, and Familial Cancer Syndrome.

In FY2010, the Bookshelf completed conversion of XML data to the NLM Book DTD format and migration to the PMC database infrastructure. Citations from GeneReviews and Essentials of Glycobiology, 2nd Edition were the first to be submitted to PubMed under the "Books and Documents" label.

The OMIM database now contains more than 20,100 records, for more than 12,500 genes and 6,000 phenotypes. The display for OMIM records was updated to mirror other redesigned Entrez databases with features such as expandable link menus. OMIM now works with the simplified URL format available in many other Entrez databases, allowing direct access to a record using the identifier (MIM number).

This past year saw the migration of the database supporting the GeneTests group from the University of Washington to NCBI. Included in that migration were the design and implementation of a public Web site, within NCBI's portal system, to support user queries and displays. NCBI implemented several functions to automate reports that GeneTests staff was generating manually, resulting not only in productivity increases for GeneTests staff, but also improved timeliness and accuracy of the information being disseminated.

Research

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

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

CBB consists of over 95 senior scientists, staff scientists, research fellows, postdoctoral fellows, and students.

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

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

The Board of Scientific Counselors (BoSC), comprised of extramural scientists, meets twice a year to review the research and development activities of NCBI and the research programs of senior investigators in the CBB. The BoSC's 34th meeting was held in April 2010.

Bioinformatics Training and Support

NCBI is responsible for: disseminating software and databases; providing support, education, and outreach to users; managing internal computing resources; and initiating and administering contracts and interagency agreements. A primary focus is on the distribution of the NCBI sequence databases using the Web, network servers, and FTP. Network (server/client) versions of BLAST and Sequin, available via FTP download, enable Internet users to run these programs locally and search the NCBI servers directly. Users with Web browsers can perform Entrez and BLAST searches, submit sequences to various databases, and obtain information about NCBI products and services. Over 98 percent of NCBI resource searches are conducted via the Web, making it the principal mode of access.

Support for users of NCBI's services has grown rapidly, which reflects the exponentially increasing number of submissions, updates, and usage, as well as increasing complexity of data due to the myriad of specialized tools produced by NCBI. A core activity has been the development of the infrastructure for procuring and maintaining high-performance computer technology to serve the ever-growing internal computing needs as well as to provide public database services. Support is also provided for the Scientific Visitor's Program, sponsorship and organization of meetings and workshops, participation in scientific exhibits, resource documentation, and publication of the NCBI newsletter. The NCBI network infrastructure was upgraded in order to support large amounts of next generation sequence data submissions. A multi-tier storage archive was implemented with disk and tape tiers to provide quick access, low-cost, high capacity storage. In addition, high-density storage systems from Panasas were installed. High-density computer infrastructure was acquired permitting 80 servers per rack to optimize compute infrastructure, reducing costs and providing additional compute infrastructure.

A Web usability team assists in improving usability of the NCBI Web site by providing guidance on better functionality, organization, automated testing, and quality control of Web resources.

Outreach and Education

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

The public services division provides user support via e-mail and telephone, staffs conference exhibits, and provides training material and seminars 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 print materials. Workshops were also provided at various scientific society annual meetings.

This year, NCBI began utilizing the social media outlets Facebook and Twitter to inform users of new

resources as well as updates and enhancements to current resources. A YouTube site was created to house videos of NCBI's 20th anniversary, GenBank's 25th anniversary, and NCBI tutorials. Tutorial titles include: "Download a custom set of records." "Obtain genomic sequence for and near a gene," and "Retrieve sequences for an Organism." NCBI provides eighteen "Announce" e-mail lists that give users the opportunity to receive information on new and updated services and resources from NCBI. RSS Web feeds are also available for updates and announcements on many of NCBI's databases.

A new Education page was launched, mirroring the newly designed Entrez pages. The redesigned page provides a collection of educational resources available on the NCBI Web site including tutorials, news and updates, documentation and help guides, community and social media sites, and courses and workshops.

Training

The NCBI Education Program was revived in FY2010 with a series of 30-60 minute webinars. These webinars are instructed by service desk staff and are broadcast live to computer classrooms at host sites. Webinar topics include: NCBI Overview, What's New at NCBI, NCBI BLAST Updates, and Genome Updates.

A series of two-day training courses titled "Discovery Workshops" was also started in FY2010. Three workshops will be held each year on the NIH campus and taught by service desk staff. The course consists of hands-on sessions that concentrate on different sets of NCBI resources: Sequences, Genomes, and Maps; Proteins, Domains, and Structures; NCBI BLAST Services; and Human Variation and Disease Genes. The first of three scheduled workshops held in September 2010 had a full registration.

EXTRAMURAL PROGRAMS

Valerie Florance, PhD
Director

The NLM Extramural Programs Division (EP) received Congressional authority for its grant programs from two different authorizing acts: the Medical Library Assistance Act (MLAA) and Public Health Law 301. The funds are expended mainly as grants-in-aid to the extramural community in support of the Library's research and training goals in informatics and knowledge management. Review and award procedures conform to NIH policies.

EP awards several categories of grants, all of which pertain to biomedical computing, informatics, and the management and dissemination of biomedical knowledge. Some resource programs, such as Grants for Scholarly Works in Biomedicine and Health, are unique to NLM. Each year, NLM makes new and/or continuing awards in every active grant category.

- Research Project grants for basic and applied research
- Resource grants for knowledge management and application of informatics
- Training and Career Development grants for informatics researchers
- Scholarly Works and Conference grants to enhance scientific and scholarly communication
- SBIR (Small Business Innovation Research) / STTR (Small Business Technology Transfer Research) grants to support informatics-related business projects

Overview of FY2010

NLM'S EP FY2010 base budget was $61,570,000 including $49,984,000 for research funding and $11,586,000 allocated to contracts for the NN/LM. That continued a trend of nearly flat budgets as the FY2009 budget was $62,121,000 comprising $49,984,000 in research funding, and $12,137,000 allocated to NN/LM contracts. Activities of the NN/LM are reported in the Library Operations section of the annual report. Highlights of this FY2010 Extramural Programs annual report:

- *Publications:* In September 2010, the NLM Board of Regents encouraged EP to begin reporting on the outputs of grant awards. Two hundred sixty-one articles were authored by NLM grantees in 2010, more than double the number published in 2007. **Table 23** provides an overview of publishing trends and hints at a trend already identified in a preliminary study – bioinformatics grantees publish more articles than their clinical informatics colleagues.
- *Web site:* To prepare for the redesigned NLM Web site, EP staff analyzed the nature and use of the EP Web site this year.
- *Award:* Dr. Bradley Malin, who received two grants from NLM in 2009, received a Presidential Early Career Award for Scientists and Engineers (PECASE) in 2010. Dr. Malin is the fourth NLM grantee to receive this honor. Earlier awardees were Dr. Russ Altman, Stanford; Dr. Ida Sim, UCSF; Dr. Kenneth Mandl, Harvard.
- *Success Rates:* A decrease in the number of applications received resulted in improved success rates for research and resource grants funded with appropriations funds.
- *ARRA:* The American Recovery and Reinvestment Act (ARRA) of 2009 continued to have a dramatic impact on EP activities in FY2010. An intensive ARRA activity of this year was providing support for HITECH grant programs of the Office of the National Coordinator for Health Information Technology, (ONC) DHHS. Support provided ranged from funding announcement development to making arrangements for non-NIH awards to flow through NIH channels, to policy guidance for grant award practices, from capture of applications from grants.gov to review and issuance of awards. While the bulk of the work was done by EP Grants Management and DEAS Extramural staff, staff from several other EP units contributed to this effort. Six EP staff received commendations from both DHHS and the NIH Director's Office for this work. A brief overview of EP's ARRA accomplishments and 2010 ARRA activities is provided in the next section.

Overview of ARRA at NLM

NIH received $10 billion from the ARRA, of which nearly $82 million was assigned to NLM. Almost 90% of that was spent on research grants and research training administered through EP. NIH received $10 billion from the ARRA, of which nearly $82 million was assigned to NLM. Almost 90% of that was spent on research grants and research training administered through EP. The rest of the funds spent on R&D contracts and other programs. All NLM ARRA funds were committed by September 30, 2010.

Table 11 summarizes the FY2010 ARRA grant awards. EP spent $37 million on ARRA grants in FY2009 and $34 million in FY2010. Over the two-year period, the NIH Office of the Director provided an additional $17,352,000 million of ARRA funds for NLM-administered grants. Some previously reviewed ARRA grants that were not funded in 2009 received funding in

2010. In FY2010, no RO1 applications were paid with ARRA funds, as the bulk of funding went to Challenge grants (commitments and new). EP's FY2010 ARRA awards were made in the following areas:

- Administrative Supplements to existing grants
- NIH Challenge Grants in Health and Science Research
- Grand Opportunities Grants
- BRDG-SPAN Small Business Pilot Program
- Small Business Catalyst Awards
- Academic Research Enhancement Award (AREA)
- Community Infrastructure Grants
- Methodology Development in Comparative
- Effectiveness Research
- Administrative Supplements for Summer Research
- Experiences for Students and Science Educators
- Restoration of approved but unfunded trainee slots to training programs
- Research & Development Contracts (issued through the NLM Contracts office)

Table 11. NLM ARRA Awards issued by EP in FY2010 – New Grants only

Mechanism of Support	Number of Awards	Amount
Challenge Grants (**RC1**)	11	$9,955,232
Community Infrastructure Grants **(RC2)**	1	$2,093,862
Development, and Growth to Spur the Acceleration of New Technologies (BRDG-SPAN) Pilot Program (**RC3**)	1	$1,627,774
Research Infrastructure Grants (**RC4**)	3	$3,709,084
Academic Research Enhancement Awards (AREA) **(R15)**	1	$429,608
Small Business Catalyst Grants **(R43)**	4	$844,780
Grand Total	**21**	**$18,660,340***

**Includes nearly $5 million in OD funds.*

Table 12. NLM ARRA Administrative Supplements issued by EP in FY2010

Actv	No.	Amount
K99	1	$ 50,000
P41	1	$ 63,097
R01	11	$ 1,096,796
R03	4	$ 92,301
R21	2	$ 172,657
T15	1	$ 95,000
Total	**20***	**$ 1,569,851**

** Includes 6 Summer Research Experiences for Students and Science Educators*

Table 13. NLM Research Funding Supported Through ARRA (FY2009-FY2010)

ARRA Awards by Location and Type of Trainees	2009	2010
US States where NLM ARRA awardees are located*	31	26
Number of Summer Research Experience trainees	67	31
Number of ARRA Informatics trainees	44	49

**See Tables 28& 29*

As part of the NIH-wide ARRA strategy, each Institute was asked to identify one or more signature areas, where $10 million or more of ARRA funds were spent. NLM identified two signature areas: Enhancing Electronic Health Records (the primary area) and Training Tomorrow's Informatics Researchers/Health IT Leaders. In 2009, NLM awarded $10,078,429 for the EHR signature area and $6,263,069 for the training area. In FY2010, EP issued seven additional awards in the EHR signature area for a total of $13,757,536 of which $6,924,830 was contributed by NIH Office of the Director.

FY2010 Appropriations Budget

Planned activities relating to the regular appropriations budget in FY2010 continued with a focus on sustaining success rates for NLM applicants and meeting NIH targets for Early Stage Investigators and new investigators.

In 2010, NIH refined target R01 grant commitment levels for new investigators by designating Early Stage Investigators (ESI) (those within 10 years of their terminal degree) as a special focus area. Two NIH targets for new investigator awards were set: (1) the success rate for new investigators obtaining R01 grants should be comparable to that of seasoned investigators, and (2) ESI should represent about 60 percent of the new investigator total. NLM met its targets for both goals through a mix of appropriations - and ARRA funds.

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 14** shows success rates in 2009 and 2010 for NLM's core grant programs for applications funded with appropriated funds.

Table 14. Success Rates - Core NLM Grant Programs, FY2009 and FY2010

Grant Program	2009	2010	% Change
Research Project Grants (R01)	17%	20%	+3%
SBIR/STTR Small Business Grants	8%	11%	+3%
Exploratory Research (R21)	9%	27%	+18%
Applied Informatics Resource Grants (G08)	10%	15%	+5%
Scholarly Works Grants (G13)	14%	20%	+6%

Research Support for Biomedical Informatics and Bioinformatics

Extramural research support is provided through a variety of grant programs that fund investigator-initiated research. EP's research grants support both basic and applied projects involving the application of computer and information science approaches in clinical medicine, translational science, public health and basic biomedical research. EP's grant program trends and accomplishments for FY2010 are summarized below for appropriated funds.

Research Grant Program

The R01 research grant program at EP receives applications through two standing program announcements with multiple deadlines. NLM's areas of interest in biomedical informatics continue to be broad. Informatics applications were received in such areas as clinical care; health services research; public health; bioinformatics; visualization; integration; mining and sharing of published knowledge or data-banked information; and translational science linking biological research findings with clinical phenotypes. Again, in 2010, NLM participated in EUREKA, with a single topic: intelligent search tool for answering clinical and biomedical research questions

- 85 reviewed R01 applications (95 in FY2009)
- 17 awarded R01 applications (16 in FY2009)

Exploratory/Developmental Grants

EP's R21 exploratory/developmental grant supports high-risk/high-yield projects, proof of concept, and work in new interdisciplinary areas. In addition to its standing program offering, in FY2010, NLM also participated in NIH-wide R21 funding opportunities on methods and approaches for detection of gene-environment interactions in human disease and health literacy.

- 26 reviewed R21 applications (32 in FY2009)
- 7 awarded R21 applications (3 in FY2009)

Resource Grants for Biomedical Informatics/ Bioinformatics

NLM's P41 Biomedical Informatics Research Resource grant program was closed in FY2009 in order to focus grant funds on innovative research in biomedical informatics rather than on research infrastructure. Limited

continuation and bridge fund arrangements were made with the five affected grantees.

Conference Grants

Support for conferences and workshops (R13) is provided through the NIH program announcement. NLM provides relatively small amounts to scientific meetings in focused areas of biomedical informatics and bioinformatics. Applicants must obtain approval from EP program staff before they can apply.

- 4 reviewed conference grants applications (2 in FY2009)
- 1 awarded conference grant application (2 in FY2009)

Small Business (SBIR/STTR)

Every year, all NIH Institutes allocate a fixed set-aside of available research funds to Small Business Innovation Research (SBIR) grants (2.5% of research grants budget), and Small Business Technology Transfer Research (STTR) grants (.3% of research grants budget). These projects may involve a Phase I grant for product design, as well as a Phase II grant for testing and prototyping. SBIR and STTR applications are reviewed by CSR.

- 24 reviewed small business applications (37 in FY2009)
- 5 awarded small business applications (3 in FY2009)

Resource Grants

Resource Grants support access to information, connect computer and communications systems, and promote collaboration in networking, integrating, and managing health-related information. The Applied Informatics series centers on optimizing the management of health-related information. These grants are not research grants and are reviewed with relevant criteria. The Scholarly Works grant program supports the preparation of scholarly manuscripts in health sciences and health public policy areas. Accomplishments in FY2010 are summarized below for appropriated funds.

Applied Informatics Grants (G08)

Last year, the open Knowledge Management/Applied Informatics grant program was suspended, and a single RFA for applied informatics grants was issued. In 2010, EP issued an RFA in this program entitled "Information Resource Grants to Reduce Health Disparities." Awards will be made in FY2011. EP anticipates issuing one topic-focused call for applications each year in this program.

- 27 reviewed Applied Informatics applications (31 in FY2009)
- 4 awarded Applied Informatics applications (3 in FY2009)

Grants for Scholarly Works (G13)

NLM alone among the Institutes is authorized to support book publications, and the Scholarly Works program continues to play a key role in important areas of biomedical scholarship, particularly in the history of science and medicine. In FY2009, the grant program moved to one application receipt date each year, with an initial two-year funding opportunity announcement (FOA). NLM plans to issue a new funding announcement in December 2010.

- 35 reviewed Scholarly Works applications (59 in FY2009)
- 7 awarded Scholarly Works applications (8 in FY2009)

Training and Career Awards

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

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

Five-year institutional training grants support pre-doctoral, post-doctoral, and short-term informatics research trainees at 18 university-based programs across the country (see **Table 14**). This program is re-competed every five years. The latest applications were received in March 2006, and five-year awards were made in FY2007. In spring 2011, NLM will issue a request for applications for the next competition.

Collectively, the programs emphasize training in health care informatics (14 programs); bioinformatics and computational biology (14 programs); clinical research translational informatics (13 programs); and public health informatics (10 programs). EP receives co-funding from the National Institute of Dental and Craniofacial Research (NIDCR) for two training slots in dental informatics at the University of Pittsburgh.

In 2005, NLM/EP and the Robert Wood Johnson Foundation (RWJF) formed a partnership to lend increased emphasis to training in public health informatics. In total, NLM received $2.89 million from RWJF through the Foundation for NIH. Four existing NLM training sites received supplemental awards to develop formal training tracks in public health informatics and to support trainees in these tracks. The four selected sites were Columbia, Johns Hopkins, Utah, and Washington. Trainees in this initiative meet twice each year for special "cohort" experiences supported by RWJF.

FY2009 was the final year in this four-year collaboration for three of four programs. In FY2010, the fourth program completed its planned activities. In total, during the funded period, 19 pre-doctoral or post-doctoral public health informatics trainees were funded through this joint NLM/RWJF program. The trainees reported their research in dozens of papers and presentations at national meetings. There is now a solid 'center' for public health informatics both inside and outside NLM's grantee community, and there are multiple venues (including AMIA, PHIN, the NLM Annual Training Conference) where public health informatics research is reported.

In 2007, NLM restructured its Short-Term Trainee Program (STTP) to focus the program on awards for minority, disabled or disadvantaged trainees, with the long-term goal of increasing diversity in the informatics research workforce. Eighteen STTP slots were awarded for use in FY2010. Programs may appoint STTP trainees at any time during the training year, which runs July 1, 2010 through June 30, 2011.

Two-year ARRA slots, with emphasis in clinical and public health informatics, were awarded to 10 programs. ARRA-funded slots supported 28 pre-doctoral and 16 post-doctoral trainees in 2009/10, and 29 pre-doctoral and 20 post-doctoral trainees in 2010/11.

Table 15. T15 Training Slots funded by NLM in FY2010 (for a total of 245.5 full-time equivalent trainees)

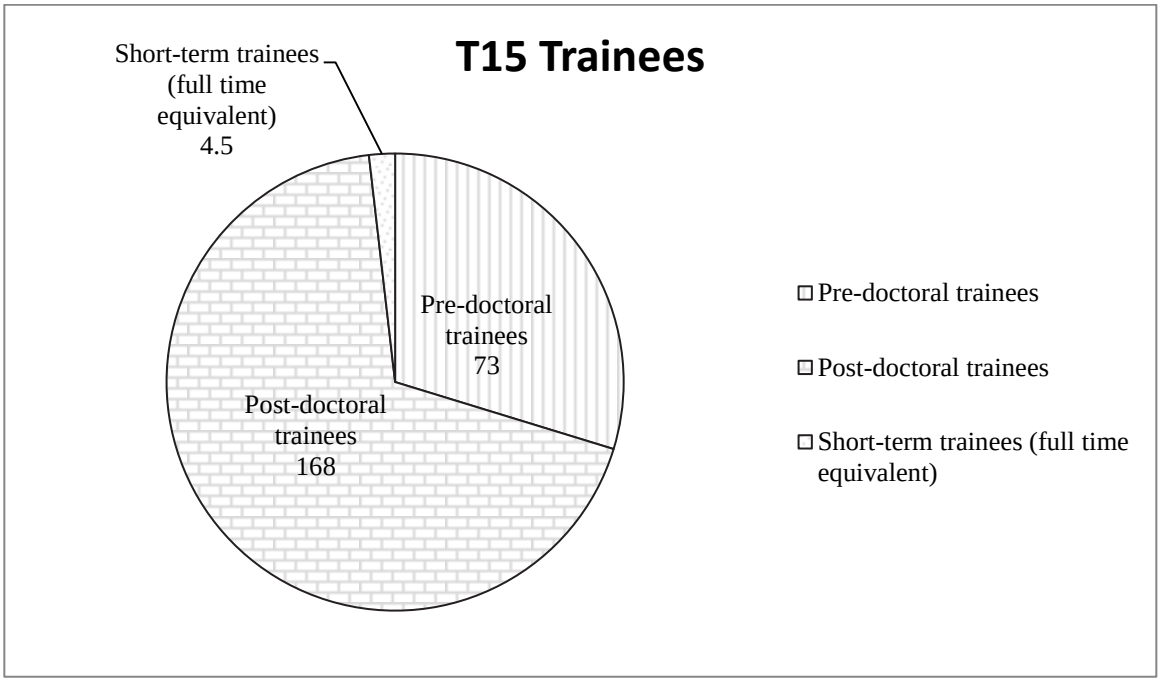
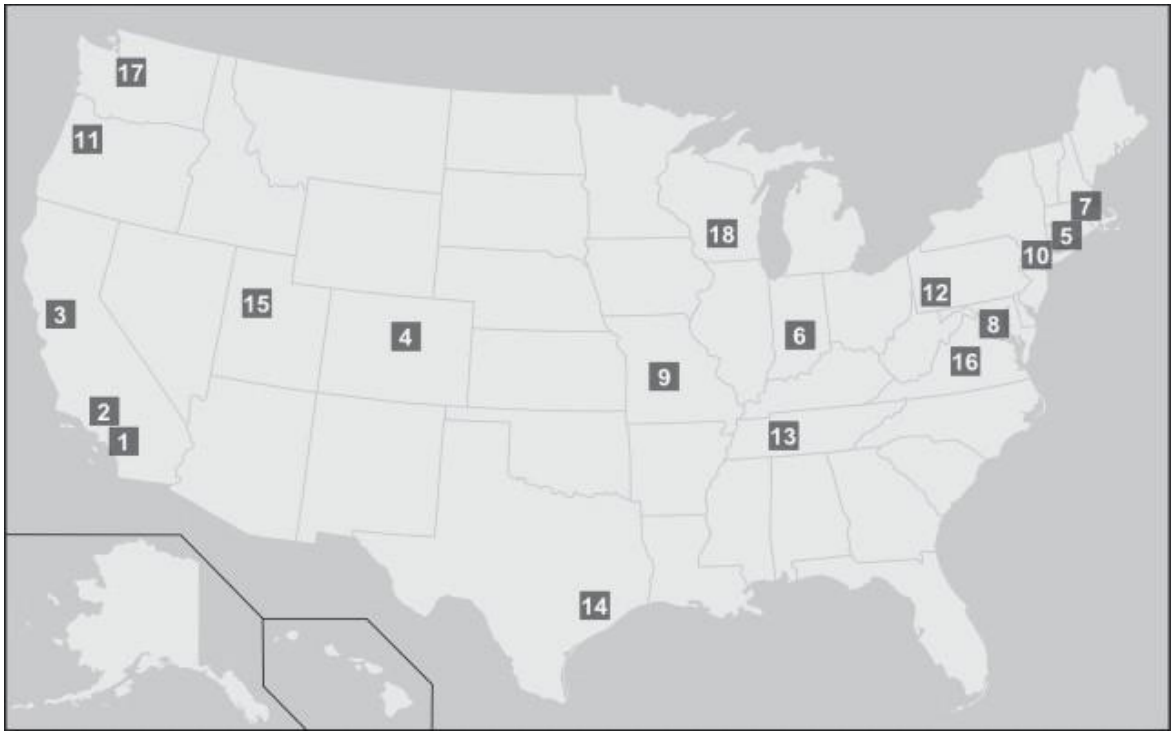


Table 16. In FY2010, NLM Sponsored Trainees at 18 Institutions



1. University of California Irvine (Irvine, CA)
2. University of California Los Angeles (Los Angeles, CA)
3. Stanford University (Stanford, CA)
4. University of Colorado Denver/HSC Aurora (Aurora, CO)
5. Yale University (New Haven, CT)
6. Indiana University - Purdue University at Indianapolis (Indianapolis, IN)
7. Harvard University (Medical School) (Boston, MA)
8. Johns Hopkins University (Baltimore, MD)
9. University of Missouri-Columbia (Columbia, MO)
10. Columbia University Health Sciences (New York, NY)
11. Oregon Health & Science University (Portland, OR)
12. University of Pittsburgh at Pittsburgh (Pittsburgh, PA)
13. Vanderbilt University (Nashville, TN)
14. Rice University (Houston, TX)
15. University of Utah (Salt Lake City, UT)
16. University of Virginia (Charlottesville, VA)
17. University of Washington (Seattle, WA)
18. University of Wisconsin Madison (Madison, WI)

Every summer, all NLM-supported trainees attend a national informatics training conference sponsored by NLM. On June 15-16, 2010, the two-day meeting took place at the University of Colorado in Denver. Research projects were presented in plenary and semi-plenary sessions by 38 informatics trainees. An additional 40 trainees presented posters at the meeting. There were 380 attendees, including directors, faculty, and staff from all NLM-funded informatics training programs; faculty and trainees from the Veteran's Administration informatics training sites; and NLM staff and guests. Two trainees received awards for the best poster, selected by their peers: Daniel Clemens of UC-Irvine, for "A Molecular Dynamics Study of the AQPO-Cam Interaction" and Michael Salmans of UC-Irvine, for "Co-factor of LIM Transcriptional Regulation in Mammary Gland Development." Two trainees received awards for the best presentation, selected by their peers: Marina Sirota of Stanford, "Discovery and Validation of Novel Drug Indications using Public Gene Expression Data" and Kenny Daily of UC-Irvine, "Genome-wide Maps of Candidate Regulatory Motif Sites."

Career Support

K99/R00 Pathway to Independence

In January 2006, NIH announced a new career transition program, the NIH Pathway to Independence (PI) award (K99/R00), which combines a two-year mentored period with a three-year un-mentored research period (the latter being similar to NLM's former K22 program). Although applications to this program are not restricted to NLM's informatics trainees, they are preferred applicants.

- 3 reviewed K99/R00 applications (6 in FY2009)
- 3 awarded K99/R00 applications (4 in FY2009)

K22 Independent Career Development Award

In FY2010, NLM re-opened its popular K22 early career award program. Several years of experience showed that the majority 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. The first awards will be issued in FY2011.

Loan Repayment Program (L30)

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

- 11 reviewed Loan Repayment Program applications (12 in FY2009)
- 4 awarded Loan Repayment Program applications (4 in FY2009)

Trans-NIH Projects

ARRA

The NIH strategy for ARRA involved EP participation in several new grant program initiatives. New topical foci included Comparative Evaluation Research, Community-based Research, Director's Opportunity Grants, and a new NIH initiative called OPPNet, designed to expand NIH's presence in social and behavioral science basic research.

NCBC

The National Centers for Biomedical Computing (NCBC) program is funded through NIH Roadmap funds under renewable cooperative agreements. An RFA for an open competition for the second five-year period of the NCBC was released on November 10, 2009. Eleven NIH ICs participated in the renewal RFA, including NLM. Twenty-one applications were received and reviewed by CSR in June 2010. Three existing and two new centers will be funded. EP administers one NCBC center, "Informatics Integrating the Bench and Bedside (i2b2)," based at Harvard's Brigham and Women's Hospital. The i2b2 NCBC will continue for the second round. The Common Fund provides full funds for i2b2 in FY2010, In future years, EP will provide an increasing share of the support for i2b2. By the final year, EP will provide 100% of the funding for i2b2. EP will also provide $50,000 co-funding to a new NCBC center named "iDASH" at University of California San Diego in FY2011 and through FY2014. EP program officers have scientific advisory roles in five NCBC centers: i2b2 at Harvard, MAGnet at Columbia University, NCIBI at University of Michigan, SIMBIOS at Stanford, and iDASH at University of California San Diego.

Multi-institute Grant Programs

NLM participates in two types of multi-institute grant programs: general and topical. General programs such as the AREA grants, diversity and reentry supplements are fundamental components of NLM's overall grant program. EP participation is confined to topical programs which do not duplicate its existing grant programs. The active multi-institute programs NLM participates in are listed in **Table 22**. The applications for multi-institute programs are reviewed by CSR, and participating institutes select grants for full or shared funding. Applications assigned to NLM that receive fundable merit scores are included in the listing for payline decisions. Links to the multi-institute initiatives in which EP participates are incorporated into the grant programs list on the EP Web site at http://www.nlm.nih.gov/ep/Grants.html.

Shared Funding for Research & Training

In FY2010, EP contributed approximately $826,051 in collaborative co-funding agreements to other NIH Institutes and Centers. The listing below outlines the projects and amount of support NLM contributed to other ICs.

- Co-funding to the National Institute of Environmental Health Sciences (NIEHS) for a grant entitled "Comparative Toxicogenomics Database (CTD)" ($196,000)
- Co-funding to the Fogarty International Center (TW) for International Informatics capacity building, for a training award to OHSU entitled "Transitional Informatics for Global Health" ($230,051)
- Co-funding the Office of the Director (OD), NIH for a grant entitled, "A Gene-Complete Computational Model of Yeast" (400,000)

In FY2010, the NLM grants which received co-funding of approximately $4,383,715 from other NIH agencies included the following:

- The NLM received funding in support of two pre-doctoral Dental Informatics trainees to the Pittsburgh Biomedical Informatics Training Program from the National Institute of Dental and Craniofacial Research (NIDCR) ($101,341)
- The OD/NIH provided full funding support for the competitive renewal of the National Center for Biocomputing (NCBC) entitled, "Informatics for Integrating Biology and the Bedside (i2b2) ($3,966,606)
- The National Institute of Biomedical Imaging and Bioengineering (NIBIB) supported a NLM conference grant for a workshop entitled, "Life Science Systems and Applications Workshop" ($3,000). The primary aim of the workshop is to provide a forum for presenting new systems advancements in emerging life science applications
- The NLM received full funding support from the NIH OD for one linked research proposal at the UCLA Consortium for Neuropsychiatric Phenomics, an NIH Roadmap Interdisciplinary Center awardee. The research grant is entitled "Hypothesis Web Development for Neuropsychiatric Phenomics," ($312,768)

Interagency Agreements and Special Initiatives

- NLM continues to provide co-funding to the NSF for the Protein Databank at Rutgers University ($200,000). This databank supplies three-dimensional representations of proteins and is the single worldwide archive of structural data for biological macromolecules.
- NLM received a final allocation of gift funding support from the Robert Wood Johnson Foundation (RWJF) through the Foundation for the NIH (FNIH) to Johns Hopkins University for a transition support supplement.

Extramural Programs Web Site
(http://www.nlm.nih.gov/ep/funded.html)

The EP Web site is an organization of cross-linked web pages and documents within the NLM Main Web site. In 2010, the EP Web site expanded its content and functionality while continuing to serve its primary customer base of researchers and students.

The EP site has expanded to 155 pages of content devoted to supporting the extramural programs. Tools such as the Research Portfolio Online Reporting Tools (RePORTer) widget have been added to enable researchers to access a wealth of information on NLM grants from 1997 to the present. In addition to increasing content, the EP Web site experienced a significant increase in usage in the past year.

There were 62,502 visitors to the site, an increase in 4% over the previous year. However these visitors viewed significantly more content than in previous years. They viewed 259,525 pages, a 40% increase, and these visitors increased the average number of pages viewed per visit by 25% (2.85 average page views per visit). In addition, the percentage of repeat visitors increased by 17%.

The EP Web site continued to have a substantial interest by international visitors who represented almost 12,000 visitors (19% of total). This increase in Web site traffic was most likely associated with interest in the ARRA awards and the changes in the NIH application and review procedures. This was supported by evaluation of the statistics on individual pages which showed that the third and fifth most popular pages were FAQ pages related to application procedures and the fourth most popular page was a tutorial on application procedures. The most popular pages were the entry and grants page.

Table 17. NLM Extramural Programs 2010 Web Trends

Factor	Value	% Change from 2009
Number of Visitors	**62,502**	**+4%**
Page Views	**259,525**	**+40%**
Avg. Pages Viewed per Visit	**2.85**	**+25%**

Increased EP Web traffic was achieved while simultaneously maintaining a high level of user satisfaction according to the ACSI survey results. The overall satisfaction score for the EP Web site was a 72 out of a possible 100, which is similar to the satisfaction ratings for the NLM main site. Below is a breakdown of individual factor scores and future behaviors. Except for the Navigation rating all of the scores are in the very good range. A score of 80 or above is considered outstanding on this instrument. The analysis was conducted based on all responses to the voluntary ASCI survey completed while the visitor was viewing an EP page (n= 44).

Table 18. Summary of NLM EP ASCI Survey 2010

Elements	Score
Content	**83**
Functionality	**77**
Look and Feel	**76**
Navigation	**69**
Privacy	**79**
Search	**74**
Site Performance	**83**
Future Behaviors	
Use as Primary Resource	**79**
Would Recommend	**74**
Will Return	**83**

Subgroup analyses were conducted on the ratings for the primary EP target groups (researchers, college/graduate students, and librarians) which represented 73% of the total respondents. The data indicate high satisfaction of the performance of the EP by the by the target groups. Although the overall satisfaction of the librarians was less than the other groups, most of the important factors from content to success at finding information were rated highly by the librarians as well. Their scores were lowered by lower rating for site features such as navigation strategy.

Table 19. Top Visitor Groups for EP ACSI Survey

Top Visitor Groups for EP ACSI Survey 11/1/09 -9/30/10			
Survey Topics	**Researcher/ Scientist**	**College/ Grad Student**	**Librarian/Information Professional**
% of respondents	**48%**	**11%**	**14%**
Overall Satisfaction	**73**	**85**	**64**
Content Score	**85**	**83**	**84**
Site Performance	**83**	**86**	**79**
Recommend site	**88**	**87**	**78**
Return to site	**89**	**84**	**87**
At least partially found information you were seeking	**95**	**100**	**83**

EP Operating Units

Program Office

Grant Program Development: Program activities in FY2010 were focused on continuing the implementation of EP's ARRA activities without neglecting the implementation of program plans for the regular grant programs. Participation in ARRA required new or expanded processes for communicating with potential applicants, selecting previously reviewed applications for awards, reviewing supplement requests, and negotiating budgets. ARRA awards affected all aspects of program work, and each program staff member took on new responsibilities to ensure that awards were made in a timely manner. Standard NIH processes also had to be employed for the new ARRA awards. For example, each ARRA award had to go through validity checks for the Research Condition and Disease Categorization (RCDC) fingerprinting.

EP's planned Challenge Grant initiative was launched in FY2008 by EP's participation in NIGMS's EUREKA program calling for high-risk, high-reward innovative applications. The NLM component requested applications in the areas of computational discovery and hypothesis testing, resulting in eight EUREKA awards funded with ARRA funds in FY2009. NLM continued its participation in the FY2010 EUREKA program, this time with an updated version of one of the ARRA challenge topics relating to intelligent information retrieval and summarization. SBIR/STTR grant programs for FY2010 focused on areas of particular interest from small business including Semantic Web technologies and intelligent search agent to improve information retrieval, knowledge discovery, and management model for biological discovery and clinical decision support. The NIH timetable for transition to electronic applications has been completed for all of NLM's active grant programs.

EP issued two funding opportunity announcements (FOAs) in FY2010. The first was a reissue of NLM's prior K22 FOA, which expired in 2006. Titled "NLM Independent Career Development Award in Biomedical Informatics," the new K22 FOA is to facilitate the transition of investigators from the mentored to the independent stage of their careers. The second FOA is a new G08 RFA, titled "NLM Information Resource Grants to Reduce Health Disparities," with the goal of supporting projects to bring useful, usable health information to health disparity populations and their health care providers.

EP convened a Discovery Informatics Grantee Meeting on July 26, 2010. Principal investigators funded under the NLM EUREKA grant program attended the meeting. The objectives of the meeting were to review the progress of the funded research, to ensure that NLM has a coherent view of the advances in the field, to provide an opportunity for collective problem-solving among the investigators, and to obtain recommendations for future research directions.

Grants Management Office

The EP/NLM grants office provides the daily and end-of-fiscal year accounting for the EP budget including all awarding mechanisms. FY2010 required additional accounting of awards for tracking grant expenditures under the ARRA funding initiative and the NIH Loan Repayment Program. Grants staff also provided support for the final year of the NLM/RWJF Public Health

Informatics Training Program. NLM grants staff is responsible for the annual maintenance of the Conflict of Interest (COI) database and the Freedom of Information (FOI) database.

NLM Grants Management Services to Other Agencies: The NLM continued its support in FY2010 of the Office of the National Coordinator (ONC) for Health Information Technology, OS, HHS. In FY2010, NLM grants and extramural staff established the grants management architecture for the program implementation and coordinated activities with the ONC related to issuance of announcements, receipt of applications and establishment of Objective Review Committees and release of awards for three HITECH grant programs funded by ARRA, totaling $48 million. Extremely aggressive deadlines required almost full-time attention to this work for several months. In addition, NLM grants management staff continued to process all noncompeting renewal cooperative agreements for awards in support the State-Level Health Information Exchange project; the Annual American Hospital Association (AHA) IT Survey, the National Health Information Network and the coordinating oversight committee for the NHIN, the National e-Health Collaborative (N***e***HC).

Scientific Review Office

Grant Review Activities: Overall, 361 applications were reviewed for which NLM was the primary assignment. Of those, 231 were reviewed by NLM. The remaining 130 applications, including all ARRA RC4/U54 Challenge grants, were reviewed by CSR. Most of NLM's non-ARRA grants reviewed by CSR were SBIR/STTR grant applications. Of the applications reviewed by NLM, 56% were in one of the four research grant mechanisms (96 R01s, 27 R21s, and 7 R13s). Knowledge management G08 grants represented 12% of the applications reviewed (27 apps.), Career Transitions Awards (K99s) – 3% (six apps.), and Scholarly Works at 25% (57 apps.) of the total applications reviewed. Other mechanisms reviewed included Loan Repayment Program applications (LRPs, 11).

BLIRC: NLM's standing review group, the Biomedical Library and Informatics Review Committee (BLIRC), evaluates grant applications assigned to NLM for possible funding. BLIRC met three times in FY2010 and reviewed 116 applications (as compared to 127 in 2009). The Committee (Appendix 1) reviews applications for most biomedical informatics and bioinformatics research applications, knowledge management/applied informatics, career support, and fellowships.

Special Emphasis Panels (SEPs): Six Special Emphasis Panels were held during FY2010 compared to 9 in FY2009. 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. NLM's SEP panels reviewed a total of 115 applications during FY2010, compared to 117 in FY2009.

This past year, we offered to help the DHHS Office of the National Coordinator for Health Information Technology (ONC) in conducting the review of applications that came in response to two of their program announcements, both related to ARRA-specified funding opportunities. They received a total of 67 applications for the "University Based Training" program initiative, all using the NIH's T15 training mechanism. The NLM/EP responsibilities were to help with reviewer recruitment, providing reviewers with the application and review guideline materials, conducting the peer review meetings/teleconferences, compiling the scores and written critiques obtained from the reviewers, and preparing summary statements. A total of 34 reviewers were recruited, which were assigned to two separate Special Emphasis Panels (SEPs), managed by the two assigned SROs (Dr. Petrosian of NLM/EP and Dr. Wiethorn of NINDS/EP). The reviews were conducted in main part following the NIH's review format, policies, and procedures. The NLM's assigned DEAS staff provided help and assistance on a number of tasks throughout the application receipt and summary statement production process. The reviewer/application assignments and the release of all summary statements were successfully completed on an exceptionally expedited schedule, within a month from the receipt of the applications.

BOR/EP Subcommittee: A second-level peer review of applications is performed by the Board of Regents. One of the Board's subcommittees, the Extramural Programs Subcommittee, conducts early concurrence reviews electronically on lists of "special" grant applications. Examples include applications for which the recommended amount of financial support is larger than some predetermined amount, or those with a high program priority but a borderline score. In 2010, the subcommittee held three early concurrence panels, including one covering ARRA Challenge Grants in the following areas: Behavior, Behavioral Change and Prevention; Comparative Effectiveness Research; Enabling Technologies; Enhancing Clinical Trials; Information Technology for Processing Health Care Data; Translational Science. The Board of Regents conducts an *en bloc* vote for all other applications that meet criteria for further consideration for funding. The EP subcommittee also reviews rebuttal requests from applicants who challenge the review of their grants. There were no rebuttals received in 2010. There were no applications referred by BOR for re-review.

Table 20. Extramural Programs Grants Budget by NIH Mechanism Groupings, and by Activity Code - FY2010

(dollars in thousands)

FY2010 Operating Budget Request by NIH Mechanism Groupings	No.	Amount
Research Project Grants (R01, R03, R21, R00, RL1, U01, DP1)	84	28,097,305
SBIR/STTR (R41, R42, R43, R44)	7	832,241
Other Research - Research Careers (K99)	4	392,465
Other Research - Biomedical Research Support (P41)	5	3,097,560
Other Research - Other (G08, G13, R13, R24, D43)	33	2,833,884
Training - Institutional (T15)	37	14,321,396
R&D Contracts (L30, L40, N01, Y01)	15	409,149
EP budget excluding TAPS and RMS/Direct Operations	**185**	**49,984,000**

(dollars in thousands)

FY2010 Operating Budget Request by Activity Code	No.	Amount
D43: International Training Grants in Epidemiology (cofund)	1	230,051
DP1:	1	400,000
G08: Knowledge Management & Applied Informatics; Planning Grant	11	1,430,557
G13: Scholarly Works in Biomedicine and Health	18	1,108,804
K99: Pathway to Independence	4	392,465
L30: Extramural Loan Repayment Program	4	199,149
L40: Extramural Loan Repayment Program	-	-
N01: NN/LM Contracts	9	-
P41: Biomedical Resource Grant	5	3,097,560
R00: Pathway to Independence	8	1,972,753
R01: Research Project Grants	65	23,844,789
R03: Small Project Grants	-	-
R13: Conference Grants	3	64,472
R15: Academic Research Enhancement Award (AREA)	-	-
R21: Exploratory/Developmental Grants	9	1,879,763
R41: Small Business Technology Transfer (STTR)	1	99,967
R42: Small Business Technology Transfer (STTR)	-	-
R43: Small Business Innovation Research (SBIR)	6	732,274
R44: Small Business Innovation Research (SBIR)	-	-
RL1: Linked Research Project Grant	1	-
T15: University Biomedical Informatics Research Training Programs	37	14,321,396
U01: Cooperative Agreement	-	-
Y01: Inter-Agency Agreement	2	210,000
EP budget excluding TAPS and RMS/Direct Operations	**185**	**49,984,000**

Table 21. FY2010 EP Grants budget (dollars in thousands)

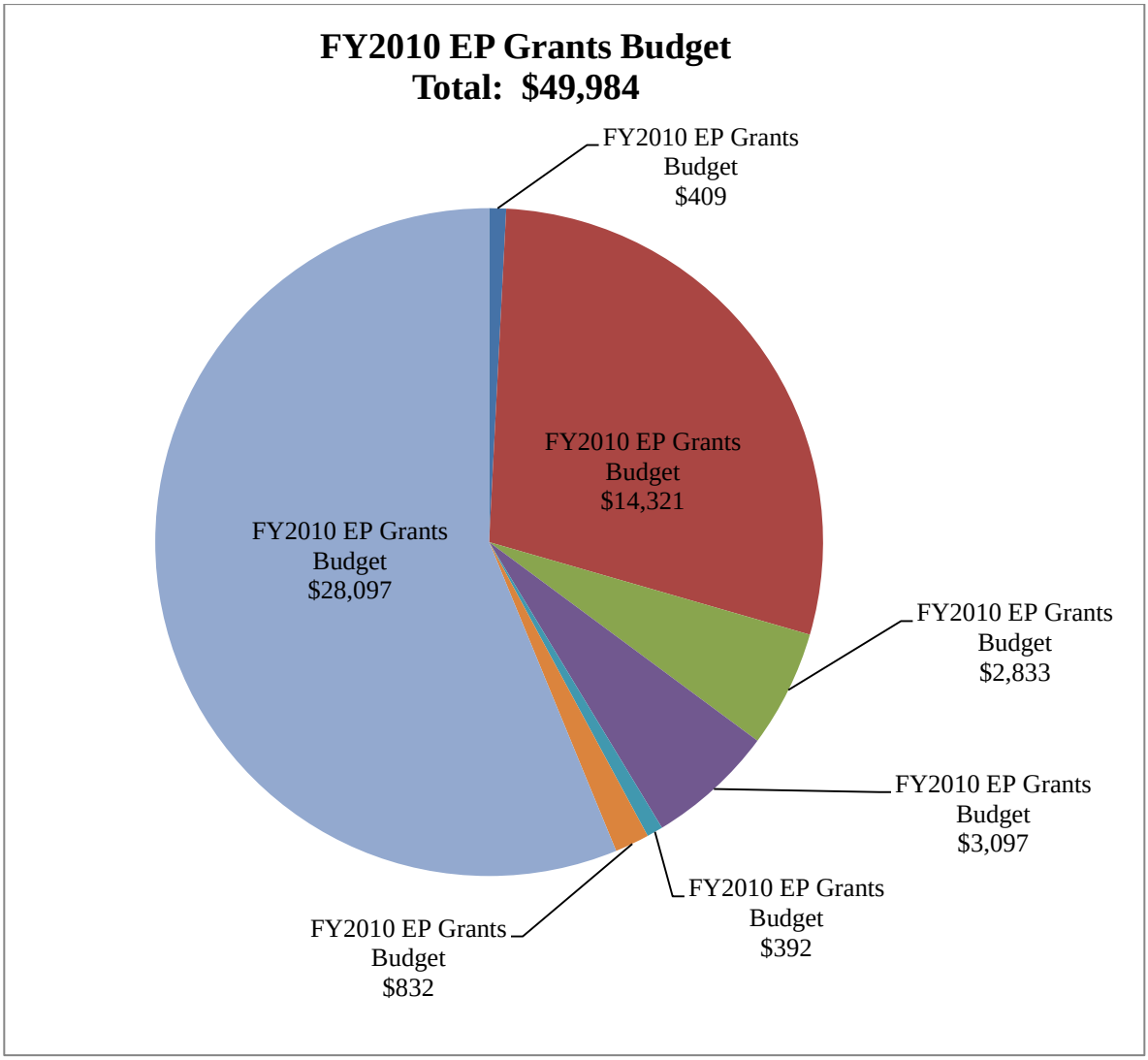


Table 22. RFA/PA Actions in FY2010

NLM's Core Active Grant Programs		
Announcement	Title	Expiration
PAR-08-080	NLM Express Research Grants in Biomedical Informatics and Bioinformatics	8-May-11
PA-10-067	Research Project Grants	8-Jan-13
PA 10-069	Exploratory/Developmental Grants (R21) for Biomedical Informatics and Bioinformatics	8-Jan-13
PA 10-071	NLM Informatics Conference Grants (R13)	8-Jan-13
PA 10-070	Academic Research Enhancement Award (AREA) (R15)	8-Jan-13
PAR 10-195	NLM Independent Career Development Award for Biomedical Informatics (K22)	8-May-13
PA-10-063	NIH Pathway to Independence (PI) Award (K99/R00)	8-Jan-13
PAR-09-030	NLM Grants for Scholarly Works in Biomedicine and Health (G13)	*Will be reissued in December 2010, with expiration date of 2/3/2012*
RFA-LM-10-001	NLM Information Resource Grants to Reduce Health Disparities (G08)	
Multi-IC Active Announcements in which NLM Participates		
Announcement	Title	Expiration
PA-08-191	Supplements to Promote Reentry into Biomedical and Behavioral Research Careers	30-Sep-11
PA-08-190	Research Supplements to Promote Diversity in Health-Related Research	30-Sep-11
PAR-10-133	Understanding and Promoting Health Literacy Research Grants (R01)	8-May-13
PAR-10-135	Understanding and Promoting Health Literacy Exploratory/Developmental Grants (R21)	8-May-13
PAR-10-136	Behavioral and Social Science Research on Understanding and Reducing Health Disparities Research Grants (R01)	12-May-13
PAR-10-137	Behavioral and Social Science Research on Understanding and Reducing Health Disparities Exploratory/Developmental Research Grants (R21)	12-May-13
PAS-10-226	Advancing Novel Science in Women's Health Research (ANSWHR) Exploratory/Developmental Research Grants (R21)	8-Jan-13
PAR 10-235	Climate Change and Health: Assessing and Modeling Population Vulnerability to Climate Change (R21)	25-May-12
PA-07-442	Extramural Loan Repayment Programs (L30)	*Ongoing; One deadline per year*
PA-07-440	Extramural Loan Repayment Programs (L40)	*Ongoing: One deadline per year*
PA-10-050	Small Business Innovation Research (SBIR) Grant Applications (R43/R44)	8-Jan-11
PA-10-051	Small Business Technology Transfer Grants (STTR) (R41/R42)	8-Jan-11

Table 23. Publications in 2010

In FY2010, NLM-funded grants led to the publication of nearly 300 articles, of which 118 were written by NLM trainees. Nearly 70 percent of the scholarly articles were written by R01 research grantees and published in academic journals ranging from *Artificial Intelligence In Medicine* to *BMC Bioinformatics* to *Journal Of Biomedicine & Biotechnology*.

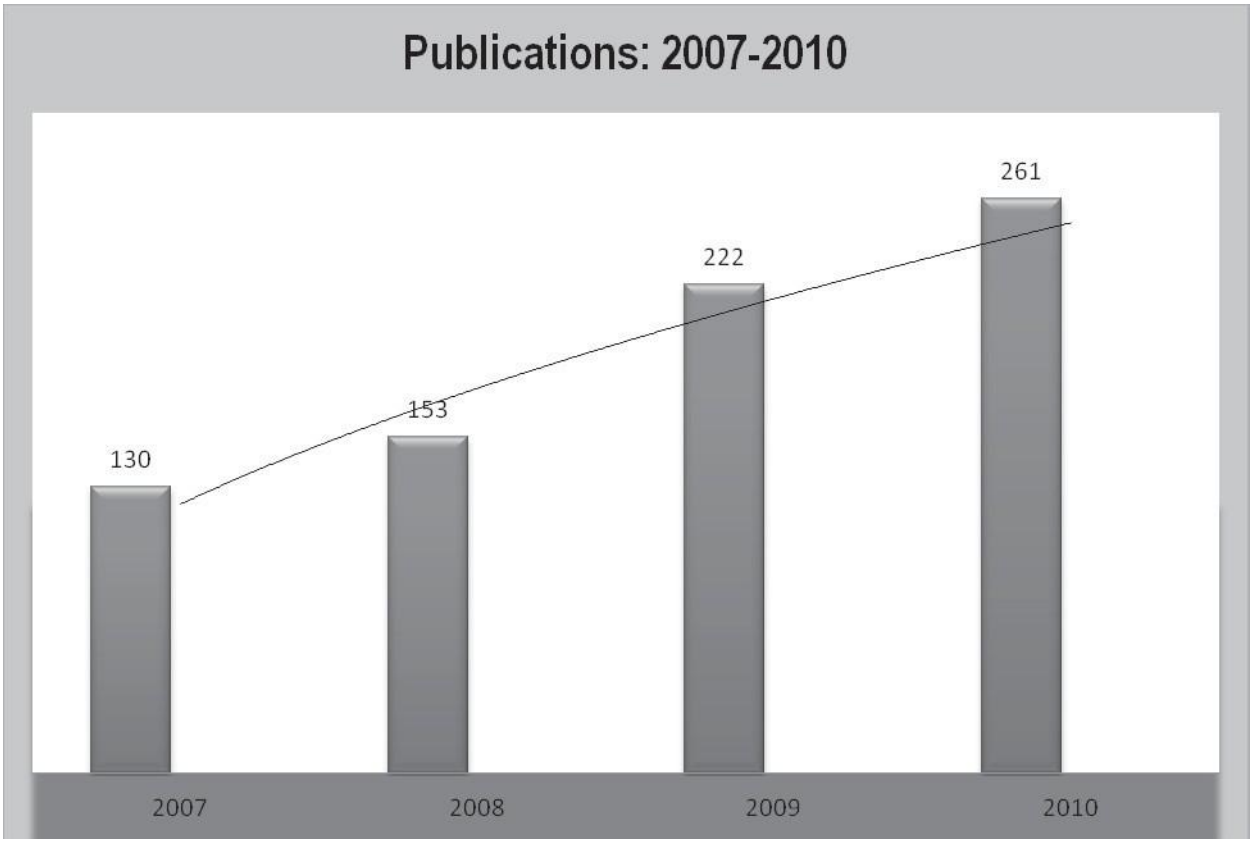


Table 24. Percentage of Increases in Published Articles of NLM-Funded Grants for 2007 – 2010

Years	Increases
2007	-
2008	**18%**
2009	**45%**
2010	**18%**
2007 to 2010	**101%**

Table 25. Top 10 Journals in which NLM Grantees published in 2010

Journal Name	No. of Articles
BMC Bioinformatics	23
Bioinformatics (Oxford, England)	20
Journal of the American Medical Informatics Association : JAMIA	15
PLoS One	9
Journal of Biomedical Informatics	8
Studies in health technology and informatics	7
IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM	6
Lecture notes in computer science	5
PLoS Computational Biology	5
Briefings in Bioinformatics	4

Table 26. Top 10 Journals in which NLM Informatics Trainees published in 2010

Journal Name	No. of Articles
Journal of the American Medical Informatics Association : JAMIA	14
Journal of Biomedical Informatics	8
Proceedings of the National Academy of Sciences of the United States of America	6
Bioinformatics (Oxford, England)	5
BMC Bioinformatics	5
Studies in health technology and informatics	5
International Journal of Medical Informatics	4
PLoS One	4
BMC Medical Genomics	3
Applied and Environmental Microbiology	2

Table 27. ARRA Funding by State: Top 10 States receiving NLM funding (FY2009 and FY2010)

State	Projects	Funding
California	28	$14,202,027
Colorado	9	$3,653,816
Massachusetts	34	$12,444,675
New York	20	$8,452,877
Oregon	8	$5,061,997
Pennsylvania	16	$8,027,998
South Carolina	8	$5,247,412
Tennessee	10	$4,471,692
Texas	11	$3,247,032
Utah	7	$4,010,025

Table 28. NLM ARRA Grant Funding by State, FY2009 (Amounts shown below are the total amounts awarded for grants in FY2009. The figures do not include funds awarded under R&D contracts.)

State	Recovery Act Projects 2009	Funding
Arizona	2	$405,916
California	17	$8,024,771
Colorado	5	$2,044,635
Connecticut	3	$243,968
Dist Of Col	1	$608,650
Florida	2	$528,873
Illinois	3	$863,789
Indiana	2	$569,921
Kentucky	1	$102,371
Louisiana	1	$209,025
Maine	2	$132,092
Maryland	4	$786,769
Massachusetts	15	$5,216,567
Michigan	2	$455,132
Minnesota	2	$514,717
Missouri	2	$579,711
Nebraska	2	$473,656
New Hampshire	2	$233,140
New Jersey	4	$698,917
New York	12	$2,515,309
North Carolina	1	$249,037
Ohio	3	$1,668,576
Oregon	4	$2,189,008
Pennsylvania	8	$4,206,910
South Carolina	5	$3,136,889
Tennessee	5	$2,434,834
Texas	8	$2,179,924
Utah	4	$1,687,238
Virginia	3	$644,946
Washington	2	$262,045
Wisconsin	4	$747,723
Total	131	$44,615,059

Table 29. NLM ARRA Grant Funding by State, FY2010 (Amounts shown below are the total amounts awarded for grants in FY2010. The figures do not include funds awarded under R&D contracts.)

State	Recovery Act Projects 2010	Funding
Arizona	1	$267,304
California	11	$6,177,256
Colorado	4	$1,609,181
Connecticut	2	$128,429
Dist Of Col	2	$1,545,200
Florida	1	$480,367
Idaho	1	$839,961
Illinois	2	$1,564,605
Indiana	2	$1,391,652
Louisiana	3	$477,104
Maryland	2	$1,364,390
Massachusetts	19	$7,228,108
Michigan	2	$604,286
Minnesota	1	$931,419
Missouri	1	$1,499,998
Nebraska	1	$377,187
New Jersey	1	$64,369
New York	8	$5,937,568
Ohio	3	$2,727,326
Oregon	4	$2,872,989
Pennsylvania	8	$3,821,088
South Carolina	3	$2,110,523
Tennessee	5	$2,036,858
Texas	3	$1,067,108
Utah	3	$2,322,787
Virginia	2	$185,103
Total	95	$49,632,166

*This does NOT include ARRA Contracts as per the data source:
http://report.nih.gov/recovery/arrafunding.aspx?mstate=&mic=LM&mfy=&mcd=&mair=false&mfun=true&msto=true&mrfa=

Table 30. NLM New Grants Awarded in FY2010
(Sorted by PI name, within each grant category)

Appropriations Budget Awards

RESEARCH GRANTS

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

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

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

HAUSKRECHT, MILOS (New Investigator)
1-R01-LM010019-01A1
UNIVERSITY OF PITTSBURGH AT PITTSBURGH
Using Medical Records Repositories to Improve the Alert System Design

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

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

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

LU, XINGHUA
1-R01-LM010144-01
MEDICAL UNIVERSITY OF SOUTH CAROLINA
Statistical Methods for Integromics Discoveries

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

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

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

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

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

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

KNOWLEDGE MANAGEMENT/APPLIED INFORMATICS GRANTS

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

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

SCHOLARLY WORKS

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

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

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

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

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

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

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

CAREER DEVELOPMENT

KALPATHY-CRAMER, JAYASHREE (New Investigator)
1-K99-LM009889-01A1
OREGON HEALTH AND SCIENCE UNIVERSITY
Clinical Image Retrieval: User Needs Assessment, Toolbox Development & Evaluation

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

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

SMALL BUSINESS GRANTS

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

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

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

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

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

Table 31. NLM New Grants Awarded in FY2010
(Sorted by PI name, within each grant category)

American Recovery and Reinvestment Act Awards

CHALLENGE GRANTS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

GRAND OPPORTUNITIES GRANTS

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

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

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

Other ARRA Research Grants (R01, R03, R21)

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

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

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

BRY, LYNN (New Investigator)
1-R01-LM010100-01
BRIGHAM AND WOMEN'S HOSPITAL
Crimson - i2b2 Integration for High-Throughput, Scalable Sample Collection

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

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

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

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

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

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

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

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

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

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

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

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

Literature and Data Driven Hypothesis Generation for High Throughput Experiments

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

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

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

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

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

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

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

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

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

VANDERBILT UNIVERSITY
Analysis Tool for Heritable and Environmental Network Associations

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

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

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

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

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

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

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

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

OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

Ivor D'Souza
Acting Director

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

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

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

Executive Summary

Enhanced MedlinePlus

MedlinePlus' greatest achievement for FY2010 was the release of the new MedlinePlus Mobile application in both English and Spanish. The mobile application is a reduced version of the full MedlinePlus Web site but contains content such as the homepage; health topics; encyclopedia; and drug information, as well as a search capability. A mobile version of the Merriam-Webster's medical dictionary was also added to the home page.

Due to decreasing usage and the availability of similar information through other online resources, NLM and its partners are phasing out the MedlinePlus Go Local project. The following Go Local sites are still available: Alabama, Indiana and Michigan.

The Medical Words Tutorial was deployed this year under MedlinePlus. The concept is to teach the layperson how to put together parts of medical terminology and acronyms.

MedlinePlus saw over 596 million page views and over 141 million unique visitors. Additionally, MedlinePlus and Spanish MedlinePlus received a score of 87 and 85 respectively on the ACSI E-Government Satisfaction Index, a survey that tracks trends in customer satisfaction.

DailyMed Project

The DailyMed project is a partnership between the Food and Drug Administration (FDA), the Veterans Administration (VA), the NLM, medication manufacturers and distributors, and healthcare information suppliers. The project seeks to provide a standard, comprehensive, up-to-date, XML-based capability for labeling the contents of medications. This year, OCCS added/updated 7,583 Structured Product Labeling (SPL) labels, published more than 2,500 unapproved drug labels and added a Universal Product Code (UPC) barcode search. In addition, DailyMed recognized more than 3 million unique visitors and increased usage by 13 percent from 62 million page views to over 70 million page views.

International Health Services Terminology Standards Development Organization (IHTSDO) Project

The International Health Services Terminology Standards Development Organization (IHTSDO) is a not-for-profit association that develops and promotes 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 FY2010, OCCS initiated several projects under the IHTSDO umbrella. These include participation in a variety of standards activities, project working groups and special interest groups, with biannual conferences and weekly conference calls; but specifically in these areas: understanding the capabilities of the IHTSDO workbench and its environment—the toolset for editing and maintaining the different products of the IHTSDO, and customizing the workbench functionality for creating and maintaining a US extension of SNOMED CT and the Unified Medical Language System (UMLS) Semantic Network.

Work has commenced to create a Request Submission and Tracking System (RQS) for users in the US to request new content or modification of existing content.

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 regarding biomedical and health-related concepts. There are currently, in the 2010AA edition of the UMLS Metathesaurus, 156 source vocabularies in 20 languages. The Metathesaurus contains over 2.2 million concepts, a 5 percent increase from last year (with nearly 9.9 million names for those concepts, a 2 percent increase), along with over 46 million relationships between and among these

concepts. This represents a nearly 2 percent growth in concepts since the 2009AB edition a year ago. In the past two release cycles (2010AA and 2010AB), more than 55 English language and 10 translation sources were successfully updated and two new sources were added in 2010AB, The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD10CM) and the Traditional Korean Medical Terms (TKMT).

A new re-engineered UMLS terminology server (UTS), along with a new licensing system has been implemented, which will replace the legacy KSS and user management systems while continuing to provide the legacy functionality.

RxNorm Project

RxNorm is a standardized nomenclature for clinical drugs which is produced by the NLM. Twelve major versions were released this year for the RxNorm Editing System. Versions 8.4 and 8.5 were released in October, 8.6 and 8.7 were released in December, 8.8 and 8.9 in February, 9.3 in March, 9.1 and 9.2 in April, 9.3 in May, 9.4 in June and 9.5 in August. Implementation of these releases included the ability to archive Branded forms, implementation of MIN term type, the institution of UMLS LVG program, and support for veterinary and human drug attributes. RxNorm currently contains 19,323 active generic drugs, 16,063 active branded drugs, 6,438 active ingredients, 13,991 active brand names and 293,227 distinct NDC codes for RxNorm forms.

Enhanced NIHSeniorHealth

With 48 topics now available in NIHSeniorHealth, many new topics were added in FY2010 including Alcohol Use and Older Adults, Life After Cancer, Creating a Family Health History, Gum (Periodontal) Disease, Exercise and Physical Activity for Older Adults, and Peripheral Arterial Disease (P.A.D.). Audio review and text-to-speech pronunciation adjustments were completed for several topics. In addition, NIHSeniorHealth recognized over 18 million site hits and over 689,000 unique visitors with over 20 percent being NIH International visitors.

Multi-layered IT Security Program

A multi-layered IT security program continued, which successfully detected more than 25.3 million probes, scans, denial of service (DOS) attacks, unauthorized access attempts and other security events─a 43 percent increase from FY2009. Major accomplishments include:

- Monitored data containing distributed denial of service (DDoS) signatures.
- Performed a monthly cycle of vulnerability scanning, detection, and remediation to improve NLM security posture.
- Performed automatic virus scanning and signature update mechanisms to combat ever increasing cyber-attacks. The Web anti-virus service has blocked over 1,500 security violations, minimizing the burden on desktop security software to fend of threats to desktop computers.
- Expanded an automatic patch management system to eliminate security vulnerabilities. Patched more than 970 desktop computers with over 8.5 million signatures, an 88 percent increase from FY2009.
- Successfully passed a network penetration test performed by an independent contractor.
- Led a successful campaign to achieve 100 percent participation in information security and privacy awareness training for NLM employees and contractors.

High Speed Communication Network

NLM improved the network to eliminate single points of failure and maintains a redundant 2 Gigabit connection to the NIHnet, managed by NIH's Center for Information Technology (CIT). In addition, NLM continues a redundant, diverse fiber connection from NLM to the MAX provided by FiberGate. It provides for increased reliability for this critical network connection. Also:

- Expanded the High Availability Computing Solution to ensure that critical applications and resources remain available to NLM users.
- Increased Citrix remote access capability for NLM flexiplace workers by 17 percent. Registration for Citrix services increased from 625 to 734 users.
- Provided network availability at more than 99.98 percent and network services availability at 99.97 percent or higher.

Data Center Reengineering

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

- Expanded data center space by an additional 2,800 square feet.
- Increased electrical power capacity by 50 percent from 900 KW to 1.35 MW.
- Increased air conditioning capacity by 33 percent, from 200 to 266 tons.
- Installed an infrastructure for 40 In-Row Coolers (IRC), which will create a high density zone that will allow power usage in excel of 12KW per cabinet.
- Installed a Bus Bar electrical distribution system reducing response time from months to minutes when increasing customer electrical power requirements.
- Installed an environmental monitoring software system that enhances the ability to manage and report environmental incidences.

- Reconfigured the data center UPS system. The facility is now supported by four 500KW UPS all in parallel with N+1 redundancy.

NLM Digital Repository Project

A video search prototype was successfully integrated into the NLM digital repository as the video player for the Films & Videos collection. This software adds innovative and unique video search capabilities to the NLM digital repository. NLM Digital Collections (http://collections.nlm.nih.gov) was launched in September as a new NLM publicly-available production application. For the Medical Heritage Library (MHL) Project**,** the History of Medicine Division (HMD) has selected nearly 6,000 NLM books that will be scanned onsite and then stored in the NLM digital repository. OCCS assisted Library Operations (LO) with technical planning for the MHL Project, including design of NLM workflows.

BHEPP

The Bethesda Hospitals' Emergency Preparedness Partnership (BHEPP) consists of the National Naval Medical Center, National Institutes of Health Clinical Center, Suburban Hospital Healthcare System, and the National Library of Medicine. OCCS led BHEPP research in the areas of: data and voice communication infrastructure to facilitate cooperation and collaboration among BHEPP member sites; disaster patient data management to investigate the practicality and effectiveness of using the digital pen technology to capture patient assessment data during triage, and disaster patient tracking to explore the effectiveness of using Real Time Location System (RTLS) technology to track disaster patients, medical responders, and critical disaster medical equipment.

Green Computing Initiatives

Green computing initiatives were advanced to more efficiently use available electrical power, cooling and computing resources. Major accomplishments include:

- Increased computer virtualization by 28 percent by implementing more virtual servers than physical servers. The increase in virtual servers leads to maximizing the utilization of available computing resources.
- Consolidated computing racks in the NLM Data Center and reduced used rack space by 25 percent.

The following describes in more detail OCCS accomplishments in FY2010:

Consumer Health

MedlinePlus

MedlinePlus' greatest achievement for FY2010 was the release of the new MedlinePlus Mobile application in both English and Spanish. The mobile application is a reduced version of the full MedlinePlus Web site but contains content such as: the homepage; health topics; and encyclopedia and drug information, as well as a search capability. A mobile version of the Merriam-Webster's medical dictionary was also added to the homepage.

There were six versions of MedlinePlus released this year: version 22.5 in November, 23.0 and 24.0 in December, 25.0 in February, 25.5 in May, and 26.0 in July. Major releases included:

- Redesign of the MedlinePlus English and Spanish public sites
- Release of a new set of Patient Instruction pages, there are currently 283 Patient Instruction pages for both English and Spanish.
- Linking between A.D.A.M. Encyclopedia articles pages and the corresponding MedlinePlus Health Topics pages and vice versa.
- Addition of over 156 Anatomy Videos in English and Spanish to the A.D.A.M. Encyclopedia

A new application was created entitled MedlinePlus Connect. The service provides a way for public entities to request the URL of MedlinePlus Topics pages that are our best match to the ICD9 code they supply. This application is being created in conjunction with the New York City-based Institute for Family Health and their MyChart Electronic Health Record and developed/implemented by a company called Epic. It expanded to over 21,000 ICD9 code mappings and 11,000 SNOMED CT mappings.

Due to decreasing usage and the availability of similar information through other online resources, NLM and its partners are phasing out the MedlinePlus Go Local project. The following Go Local sites are still available: Alabama, Indiana and Michigan.

Other major enhancements to MedlinePlus include the release of an English-only Web service that allows developers to build applications that send structured XML requests and receive structured XML responses. The MedlinePlus Vocabulary and Summaries in XML format was released, which allows the file to be available to the general public. There were also over 141 production requests implemented this year.

The Medical Words Tutorial was deployed this year under MedlinePlus. The concept is to teach the layperson how to put together parts of medical terminology and acronyms. It follows a standard tutorial slideshow format with reinforcement activities.

MedlinePlus saw over 596 million page views and over 141 million unique visitors. Additionally, MedlinePlus and Spanish MedlinePlus received a score of 87 and 85 respectively on the ACSI E-Government

Satisfaction Index, a survey that tracks trends in customer satisfaction.

DailyMed Project

The DailyMed project is a partnership between the Food and Drug Administration (FDA), the Veterans Administration (VA), the NLM, medication manufacturers and distributors, and healthcare information suppliers. The project seeks to provide a standard, comprehensive, up-to-date, XML-based capability for labeling the contents of medications. This year OCCS:

- Added/updated 7,583 Structured Product Labeling (SPL) labels.
- Published more than 2,500 unapproved drug labels.
- Added SPL Imprint Data and SPL Manufacturer data to RxNorm attributes.
- Added NDC search functionality to DailyMed.
- Added Universal Product Code (UPC) barcode search.
- Recognized more than 3 million unique visitors.
- Increased usage by 13 percent from 62 million page views to over 70 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. The system includes the Accent "Talking Web" module developed by OCCS to provide accessibility enhancements, including a selectable range of type sizes and spoken text. With 48 topics now available in NIHSeniorHealth , many new topics were added this year, including Alcohol Use and Older Adults, Life After Cancer, Creating a Family Health History, Gum (Periodontal) Disease, Exercise and Physical Activity for Older Adults, and Peripheral Arterial Disease (P.A.D.). Audio review and text-to-speech pronunciation adjustments were completed for several topics. The entire site was regenerated with global speech tuning.

The Accent module received numerous enhancements including the addition of several speech tunings to the back-end lexicon/dictionary. In addition, NIHSeniorHealth recognized over 18 million site hits and over 689,000 unique visitors with over 20 percent being International visitors.

IT Security

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

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

NLM contracted for an independent network and system assessment and penetration test. The findings resulted in network environments to be highly resilient to threats.

With the increase of new vulnerabilities and the rapid emergence of associated threats, there must be a deployment of more software patches than ever before, with a much greater degree of urgency. NLM's automated patch management program applied over 132,000 patches on commodity desktops, fixing known vulnerabilities to software.

Anti-virus screening continued for all outgoing Web (HTTP) connections to the Internet. This Web anti-virus service has blocked over 1,500 security violations, minimizing the burden on desktop security software to fend off threats to desktop computers.

A distributed denial of service attack (DDoS) occurs when multiple systems flood bandwidth or resources of a targeted system. NLM has DDoS mitigation protection through a contract with Prolexic, Inc. The NLM Internet traffic is constantly monitored for data containing potential DDoS signatures and, if any are detected, NLM can switch its data traffic to the Prolexic network for removal of the harmful DDoS traffic.

More than 970 desktop computers were updated with over 8.5 million signatures, an 88 percent increase from FY2009 OCCS handled over 25.3 million intrusion detection system alerts, an increase over the more than 14.5 million alerts in FY2009.

In order to support the NIH information security management goals, NLM collaborated with NIH on the first major HHS enterprise-wide security solution, ArcSight, HHS' Distributed Security Event and Information Management (SEIM) System for real-time detection and alerting, and long-term security log analytics. In addition, due to connectivity and Internet traffic volume, NLM played a key role in vendor and product selection for HHS' Enterprise Intrusion Detection (IDS) and Packet Analysis System.

As a member of the NIH Tiered Security Policy Compliance Workgroup, NLM collaborated in the development of the NIH Policy for Information Security Policy Development, which designates responsibility and ensures consistency in policy development, review and approval.

The Office of Management and Budget (OMB) requires that HHS computer users complete annual IT security awareness training. NLM has completed 100 percent of the mandatory FY2010 Security Awareness Training for employees, contractors and fellows. An NLM policy requires completion of the training within five days for new hires.

High Speed Communication Network

Reliable LAN and Internet communications services continued, meeting the data communications needs for new IT systems, providing security services as well as end user assistance, implementing new network-based applications and operating systems, and exploring new

technologies and plans to meet NLM's continued growth in networking and communications. Steps were taken to increase the capabilities and reliability of network services and storage by providing for the following:

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

Public Internet connectivity services to NLM are provided through a contract with Level 3. Internet connectivity was upgraded to a full Gigabit connection in September of 2009. This primary circuit, and the redundant diverse backup circuit, connects NLM to the Level 3 Internet point-of-presence in McLean, VA. CIT and NLM have a peering arrangement where, in the event that the primary and backup NLM Level 3 circuits fail, NLM Internet services will automatically failover to use the CIT Internet connections to Level 3 and Time Warner. This failover capability is tested once per month.

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

The NLM perimeter network provides a 10 Gigabit security boundary to aggregate connections to NIHnet, Internet, Internet-2, and the NCCS. It also provides for 10 Gigabit connections to interconnect OCCS, LHNCBC, and NCBI divisional networks. A one Gigabit private, dark fiber connection has been deployed from the NLM main campus site to NLM satellite sites EP, SIS, OCCS and OAM.

State-of-the-art firewalls guard the network at the perimeter and lower layers within the local network. The NLM perimeter and NCCS firewalls run on 10 Gigabit appliances.

NLM's Citrix remote access capability was increased by 17 percent. Registration for Citrix services increased from 625 to 734 users.

Network availability was provided at more than 99.98 percent and network services availability at 99.97 percent or higher.

Implementation of the High Availability Computing Solution continued to ensure that critical applications and resources remain available to NLM users. Work continued on clustered Oracle server system filers and clustered storage systems as NLM's high availability computing resources.

Data Center Reengineering

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

- Expanded data center space by an additional 2,800 square feet.
- Increased electrical power capacity by 50 percent from 900 KW to 1.35 MW.
- Increased air conditioning capacity by 33 percent from 200 to 266 tons.
- Installed an infrastructure for 40 In-Row Coolers (IRC). The IRCs will create a high density zone that will allow power usage in excess of 12 KWs per cabinet. Thirty-six of the 40 IRCs have been installed and configured in groups of eight, for a total of five groups. Each group is configured for N+1 redundancy.
- Installed additional cable trays for backbone blown-fiber optic cabling throughout Buildings 38 and 38A continued during the year. This is designed to provide alternate diverse cable paths in the event that a disaster destroys one path of the cabling.
- Installed an environmental monitoring software system within the data center that has enhanced the ability to manage and report environmental incidences. The environmental system provides real-time status of temperature, humidity, UPS power usage per phase as well as CRAC status.
- Installed a Bus Bar electrical distribution system. The Bus Bar improves electrical power distribution as well as reduces the response time from months to minutes when increasing customer electrical power requirements.
- Reconfigured the B1 data center's UPS system. The facility is now supported by four 500 KW UPS, all in parallel with N+1 redundancy.
- Increased L5-20 circuits from 336 to 353, in order to keep up with growing electrical power demands.

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. Currently, in the 2010AA edition of the UMLS Metathesaurus, there are 156 source vocabularies in 20 languages. The Metathesaurus contains over 2.2 million concepts, a 5

percent increase from last year (with nearly 9.9 million names for those concepts, a 2 percent increase), along with over 46 million relationships between and among these concepts. This represents a nearly 2 percent growth in concepts since the 2009AB edition a year ago. In the past two release cycles (2010AA and 2010AB), more than 55 English language and 10 translation sources were successfully updated. Two new sources were added in 2010AB: The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD10CM) and the Traditional Korean Medical Terms (TKMT). The SPECIALIST tools and the lexicon were also updated in the 2010AA edition.

A new set of source documentation was introduced, describing in detail how individual terminologies are represented in Rich Release Format (RRF). For each updated source, there are two new Web pages: the "Source Representation" page, which lists specific source data elements and provides information on their representation in the UMLS; and the "Metathesaurus Representation" page, which lists RRF data elements and traces them back to the specific source data that populates them.

A newly designed common technology platform for supporting UMLS applications was implemented, which acts as the underlying object model for MetamorphoSys, Unified Terminology Services (UTS) (formerly known as the KSS), and Metathesaurus Editing and Maintenance Environment (MEME). The domain model itself is of particular interest and importance, as it is a comprehensive model of all aspects of terminology as understood by the UMLS.

A new re-engineered UMLS terminology server (UTS), along with a new licensing system, has been implemented, which will replace the legacy KSS and user management systems while continuing to provide the legacy functionality. This system incorporates KSS-style content and authentication web services, new browsers for the UMLS Metathesaurus and Semantic Network, a user management application, and a newly designed Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT) browser. The UTS system is expected to be deployed and go into service later this year in a stable and robust production environment.

International Health Services Terminology Standards Development Organization (IHTSDO) Project

The International Health Services Terminology Standards Development Organization (IHTSDO) is a not-for-profit association that develops and promotes 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 FY2010, OCCS initiated several projects under the IHTSDO umbrella. These include participation in a variety of standards activities, project working groups and special interest groups, with biannual conferences and weekly conference calls, but specifically in these areas: understanding the capabilities of the IHTSDO workbench and its environment─the toolset for editing and maintaining the different products of the IHTSDO; and customizing the workbench functionality for creating and maintaining a US extension of SNOMED CT and the Unified Medical Language System (UMLS) Semantic Network.

Work has commenced to create a Request Submission and Tracking System (RQS) for users in the US to request new content or modification of existing content. The system is designed to be transparent and open, to allow users to track requests as they move through the process and provide feedback when necessary. When functional, the RQS will be integrated with the workbench itself and with other request submission systems, such as the one for the IHTSDO itself. Portions of the application will be restricted to UMLS licensees with ties to the UTS, to search for SNOMED CT content in the UTS browser.

The team worked with IHTSDO Mapping SIG, and made significant progress on a workable short-term environment for development of a SNOMED CT to ICD-10 map, initially for a priority list of SNOMED CT concepts. Using an eclipse-based tool developed by the NHS in the UK, NLM developed a comprehensive workflow for managing development, editing, and conflict resolution for these mappings. Supporting tooling was also developed to enact various administrative actions of the workflow environment. Work began on planning to integrate these workflow tools with IHTSDO workbench in the next few months.

Development began on comprehensive plans for US Extension maintenance in a recently acquired NLM namespace, including evaluation of the need for a portal Web site, documentation, tracking, workbench integration, participant authentication, etc. Work also began on creating NLM infrastructure to support US Extension maintenance which includes: building of US extension-specific workbench IDEs; public-facing NLM subversion for collaborative editing; initial planning for back-end infrastructure to support publication; and QA of first (and subsequent) editions of the US extension.

RxNorm Project

RxNorm is a standardized nomenclature for clinical drugs which is produced by the NLM. Twelve major versions were released this year for the RxNorm Editing System. Versions 8.4 and 8.5 were released in October, 8.6 and 8.7 were released in December, 8.8 and 8.9 in February, 9.3 in March, 9.1 and 9.2 in April, 9.3 in May, 9.4 in June and 9.5 in August. Implementation of these releases included the ability to archive Branded forms, implementation of MIN term type, the institution of UMLS LVG program, and support for veterinary and human drug attributes.

There were 12 monthly releases to RxNorm this year which included the regular monthly releases. OCCS

completed the monthly inversion and insertions of nine regular sources, added 50 new quality assurance checks to the database, processed 10,756 concepts for data quality purposes, and inserted 252 sources. Two resynchronizations were completed with UMLS data, and updated resynchronization code for processing PIN and MIN term types, and updated preferred term processing for new merge scenarios.

The RxNorm application currently contains 19,323 active generic drugs, 16,063 active branded drugs, 6,438 active ingredients, 13,991 active brand names and 293,227 distinct NDC codes for RxNorm forms.

Medical Subject Headings (MeSH) and Related Systems

MeSH is the National Library of Medicine's controlled vocabulary used for indexing articles for MEDLINE/PubMed. MeSH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts. Accomplishments included:

- Updated the information pages to view and compare the MeSH Translation Maintenance System (MTMS) statistical information in various years.
- Completed the MTMS cutover for the French, German, Italian, Czech and Croatian translations.
- Generated and distributed the new 2011 MeSH XML for various translations.
- Loaded the Swedish and Norwegian translations.
- Completed development of the Journal Descriptor Maintenance System (JDMS) for the Global Citation Management System (GCMS).
- Completed the current Year End Processing (YEP).
- Implemented the cutover for the 2011 MeSH.
- Generated new MeSH MARC and new automatic and manual tasks XML files for 2011 MeSH.
- Completed development and implementation of the new handling for Thesaurus Authority and Semantic Types Authority in the M2000 Editing System.
- Completed design and development of the new Automated Documentation System to create data dictionaries and program specifications for the M2010 Editor.
- Completed database design for the M2010 Editing System.
- Completed scripts to define Production and Archive databases for the new MeSH editing system.
- Completed data load to the MeSH Archive Database.
- Completed system specifications and development for the User Profile Maintenance System.
- Completed program specifications and development of the Search Module in the M2010 Editor.

Research and Development

NLM Digital Repository Project

A video search prototype was successfully integrated into the NLM digital repository as the video player for the Films & Videos collection. This software adds innovative and unique video search capabilities to the NLM digital repository. An approach was developed for creating captions and transcripts using speech recognition and text editing tools, and created Section 508-compliant captions and searchable transcripts for a set of 11 History of Medicine (HMD) videos. With the assistance of a summer intern, 28 additional videos were captioned. OCCS selected video conversion tools, generated video derivative files in several access formats, and ingested all video files and metadata into the repository. OCCS demonstrated the video search tool to the DRIG, the Director of OCCS, the Digital Repository Oversight Group, the DC Fedora Users Group, and to representatives from the National Agricultural Library.

The NLM Digital Collections was launched in September as a new NLM publicly-available production application. Traffic, system load, and application logs were monitored during the days following the launch to verify smooth operation of this new system. The initial content includes 518 cholera books and 11 historic films digitized from HMD collections. Several thousand users visited the site in the first few days after launch. Users can browse and search repository content using the repository user interface, or they can search for repository content from the NLM home page using NLM's Vivisimo search engine.

For the Medical Heritage Library (MHL) Project**,** HMD has selected nearly 6,000 NLM books that will be scanned onsite and then stored in the NLM digital repository. OCCS assisted LO with technical planning for the MHL Project, including design of NLM workflows, and review of technical specifications for the Kirtas scanner, as well as assisted the LO Preservation Section define system requirements and develop a new Linux-based QA System that will validate the completeness and data integrity of all image, metadata, and other files produced by Kirtas KQ Services, and ensure that the digitized MHL books can be successfully ingested into the NLM digital repository.

Database

OCCS migrated the RxNorm, development and QA RAC, Siebel and Relais databases to 11g, planned the architecture for release 2 of the KSS application, and built a new RAC One Node database for the UMLS Knowledge Server II project.

Search Engine

Upgraded Vivisimo to version 7.5, developed updates to MedlinePlus search for the MedlinePlus user interface redesign, released the MedlinePlus user interface redesign search, developed and released a MedlinePlus drug

information web service using Vivisimo search as well as released the HMD Collections Directory search and the HMD Finding Aids Consortium search engine.

ReportNet Migration

All applications were migrated from Impromptu Reports to COGNOS8. Work was completed on the acquisitions and cataloging packages for Voyager COGNOS users as well as updating COGNOS to version 8.4 for the development and production environments.

Comparative Effectiveness Research (CER)

This web resource was developed to help inform investigators of comparative effectiveness. Comparative effectiveness research (CER) is a type of health care research that compares the results of one approach for managing a disease to the results of other approaches. This year the query page was modified and new content was added. The query pages were also redesigned to allow for the same formatting as the new PubMed search page.

Web Content Management (TeamSite)

NLM uses TeamSite to provide content and application management for Web sites. TeamSite was upgraded to version 6.7 in October, which included a new WYSIWYG editor, TinyMCE, as well as significant changes in user authentication and authorization, providing more flexible role management. The hardware was upgraded as well as the software to current revisions (Solaris 10, Samba 3.4, Apache 2.2, etc). SiteMinder was integrated with TeamSite to provide single sign-on to the application. This also prepares TeamSite to support PIV authentication when the Center for Information Technology (CIT) reconfigures SiteMinder.

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. The application was upgraded from version 8.1a to 8.5a in April then from 8.5a to 8.7d in May.

Web Support

The NLM Emergency Access Initiative site was quickly launched in response to the Haiti earthquake. EAI provides access to pertinent online books and journal articles to health care professional's librarians and the public affected by disasters, which may not otherwise have access to this information due to loss of local resources.

The *NLM in Focus* newsletter was launched using a MovableType blogging platform. *NLM in Focus* is the new online newsletter produced by the Office of Communications and Public Liaison (OCPL). OCCS worked with the Library Operation's (LO) Public Support Division (PSD) Reference and Web Services Section (RWS) Web team and NLM OCPL to provide an easy-to-use content management platform using MovableType so that OCPL can quickly publish new articles, keeping customers up-to-date on NLM's rapid changes.

Bethesda Hospital Partnership

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

OCCS led BHEPP research in the following areas:

- Data & voice communication infrastructure─to facilitate cooperation and collaboration among BHEPP member sites by providing a laser communications network and linking the disparate voice communications systems into a unified interoperable system.
- Disaster patient data management which contains Patient Data Capture and Patient Data Exchange systems - The Patient Data Capture using the digital pen technology to capture patient assessment data during triage. The Patient Data Exchange exchanging a minimum core patient data among BHEPP member sites to support disaster response operations.
- Disaster patient tracking─to explore the effectiveness of using the Real Time Location System (RTLS) technology to track disaster patients, medical responders, and critical disaster medical equipments within a facility/hospital and between facilities/hospitals.

Green Computing Initiatives

Green computing initiatives were advanced to more efficiently use available electrical power, cooling and computing resources. Major accomplishments include:

- Increased computer virtualization by 28 percent by implementing more virtual servers than physical servers. The increase in virtual servers leads to maximizing the utilization of available computing resources. In FY2008, the ratio of virtual servers to physical servers was 91:309, and 123:214 in FY2009. In FY2010, the ratio of virtual servers to physical servers changed to 231:274, a 46 percent virtualization.
- Consolidated computing racks in the NLM Data Center and reduced used rack space by 25 percent.

The freed space was made available to support other growth areas within NLM.

- Upgraded older equipment to more energy efficient hardware.
- Plugged gaps within cabinets to redirect cold air for equipment cooling.

Desktop virtualization was introduced with virtual machines (VM) being integrated into desktop operations. Use of a VM allows for flexibility allowing the user to access the VM from their primary computer without requiring them to maintain a second physical computer. In addition, several software applications were virtualized such as UMLS Terminal Services (UTS), Cognos, Digital Library Extension Service (DLXS), and Vivisimo. Software virtualization benefits include using fewer resources, accelerated application deployment, and implementation of security by removing the requirement for end-users to have administrator privileges. Desktop virtualization is currently supporting over 500 NLM users.

Medical Literature Support & Document Delivery Services

Data Creation and Maintenance System (DCMS)

The major event for the Data Creation and Maintenance System (DCMS) this year is the baseline extraction, which is a re-release of all DCMS citations that follows the MeSH Year-end Processing (YEP). The 2010 MEDLINE®/PubMed® baseline database contains 18.5 million records. OCCS loaded and processed over 63,000 "new" OLDMEDLINE records for the publisher year 1947, and completed mapping over 1.4 million of the more than 1.8 million records of OLDMEDLINE terms to current MeSH terms and switched the status to Medline. Also:

- Reduced the elapsed time for extraction of the baseline to five hours and 57 minutes, a 45 percent reduction over the time required in the prior year. The number of records exported per hour increased by 89 percent.
- As part of the annual Meeting Abstracts baseline extraction, Lister Hill Center's (LHNCBC) MTI (Medical Text Indexer) indexing for more than 114,000 records were loaded and re-released to the Gateway system.
- Completed loading of the Publisher Back issue and PubMed Central data, which included 14,600 total files loaded, 194, 838 total citations loaded, and 13,519 new abstracts processed.
- A new version of the DCMS Loader & Extractor (Makexml 1.8) was installed in March.

DOCLINE

DOCLINE, the NLM interlibrary loan (ILL) system, supported approximately 3,185 domestic and international libraries in processing more than 1.8 million interlibrary loan transactions and 328,000 Loansome Doc requests this year. A new major version of DOCLINE was released on August 8, and three maintenance releases were also promoted. In FY2010, the Requests modules were redesigned for more efficient work flow in the Borrow, Lend, and Status functions. New features included addition of Odyssey as a delivery method, the ability to skip specified libraries during routing, a "finish now" button for express request placement, ability to route on electronic and/or text format, warning messages for duplicate requests, and alert messages if an erratum exists for the article being ordered. Also, emphasis was placed on an updated Web design and clearer terminology. There were 169 DOCLINE and Loansome Doc enhancement requests implemented in addition to internal security and server configuration changes. DOCLINE users ordered 926,170 distinct articles from 22,424 journals and 14,996 monographs. Also, there were over 24 million page views and over 1.1 million visitors to the site.

Voyager Integrated Library System (ILS)

The year-end processing of the Voyager Integrated Library System database was completed which included altering numerous fields and upgrading the MeSH terms to the 2010 version. Over 76,000 records were affected. Baseline files of all records were created in both MARC and XML formats. These included approximately 1.2 million catfile records, 1.4 million catplus records, 130,000 serfile records, and 790,000 authority records.

ENCompass Support

eTK (electronic Thorndike and Kibre) is a digitized version of a printed work, A Catalogue of Incipits of Mediaeval Scientific Writings in Latin by Thorndike and Kibre. eVK (electronic Voights and Kurtz) is an electronic resource called A Catalogue of Incipits of Mediaeval Scientific Writings in Old and Middle English, an electronic resource edited by Linda Voights and Patricia Kurtz. There were more than 43,000 incipits of medieval (medical) manuscripts, and data relating to them in a text file which was a database dump from the University of Missouri. The main achievement for this year was the converting of raw eTK/eVK data from the University of Missouri to XML format for loading into the ENCompass system by Ex Libris.

Relais

NLM uses the commercial off-the-shelf Relais system for electronic document delivery and Interlibrary Loan (ILL) management. Documents requested via DOCLINE are scanned and automatically delivered using the borrower's requested delivery method. A new version of Relais was installed and the changes were made in the stored

procedure to allow Odyssey delivery by NLM and the NIH library.

ScanTrac (PubMed Central Inventory)

PubMed Central (PMC) is NIH's free digital archive of biomedical and life sciences journal literature. ScanTrac tracks activity in the project to scan back issues of journals to add to PMC and now has tracking data for 560 journals (55,735 issues). Most have been entered into PMC.

Print Journal Donation Tracking System

An automatic system to track the print journal donations was implemented. This system would facilitate donation of material missing in the NLM collection from libraries, etc. using DOCLINE holdings if available. Three hundred and two donations have been submitted since the application launched, 686 donors and 606 institutions registered the system. There are 1,387 total titles offered, 3,562 total volumes offered and 1,003 titles received.

Literature Selection Technical Review Committee (LSTRC)

Several modifications were made to the MEDLINE Review application, which is used to review journals for inclusion in MEDLINE, NLM's bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences. The MEDLINE Review Application Form is required for all journals initially reviewed by the LSTRC for inclusion in MEDLINE. Enhancements were added to support Administrators in completing administrator functions, such as assigning users/reviewers, allowing reviewers to add/update their review form online and viewing completed review forms of their associates.

Serials Extract File (SEF)

Among numerous upgrades and fixes, a change request to modify algorithm for extracting publisher data from ILS (Voyager) and importing it to the Serials Extract File (SEF) was completed. There were more than 21,000 SEF records updated. The 2010 List of Serials Indexed (LSI) publication was generated in XML format for the NLM ftp server. Year-end processing (YEP) for Serials was completed.

NLM Classification System

The NLM Classification System allows public and institutional access to the NLM Classification and related services and includes a Classification Editor. Publication of printed editions ceased with the 5th revised edition in 1999. The major Classification re-engineering accomplishments this year included:

- Updating the database to get the latest changes in MeSH;
- Preparing the Classification Editor for manual updates; and
- Preparing the final updated changes and generating HTML pages and PDF documents.

Business Continuity and Disaster Recovery

In order to protect NLM's mission-critical systems the Center for Information Technology (CIT) and NLM have implemented an NIH Consolidated Colocation Site (NCCS) in Sterling, Virginia. The NCCS is operational with initial capabilities as a disaster recovery and load-balancing site. The NCCS serves as a disaster recovery/alternate computing site for NLM as well as CIT, the National Cancer Institute (NCI), National Heart Lung and Blood Institute (NHLBI), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), Office of the Director/Office of Research Services (OD/ORS) and Department of Health and Human Services/Office of the Secretary (HHS/OS).

This year, the NCCS was moved to a larger location creating a framework for a more robust cloud and disaster recovery architecture. This sets the stage to support disaster recover across all NLM application and system tiers, which previously had not been supported.

At present, all NLM mission-critical systems are either under active/active, active/passive or active/cold-backup mode depending on their business requirements. The Business Continuity and Disaster Recovery Plan covers NCCS as the primary resource for system restoration and uninterrupted processing if the primary NLM computing facilities on the NIH campus are rendered unavailable by a disaster or other contingency. VPN and Citrix remote access services are implemented at the NCCS. Various other upgrades were performed to the storage systems and servers located at this site.

Cloud Computing

Cloud computing is a technology that delivers hosted services over the internet and allows for more efficient computing by centralizing storage, memory, processing and bandwidth. This year, NLM began evaluation of cloud services and procured one year of infrastructure-as-a-service with Terremark Federal. The MedlinePLus and PHPartners web sites, as well as the Digital Library Extension Service (DLXS) search and delivery platform, were evaluated utilizing cloud services. This evaluation will continue in FY2011 with specific focus on defining security controls for a cloud implementation that assures NLM's security best practices and requirements.

Office Automation & Customer Support

A "Store" model was adopted in FY2009 to serve hardware and software to its constituent users. In past years, equipment was procured as requested by the user.

This year an OCCS Store catalog was developed listing standard items and accessories along with pricing. This serves as a quick reference purchasing guide for customers reducing time from order to installation.

Altiris software metering agent was implemented, which allows for discovery of applications on all work stations via the metering process. Software delivery limited licensed collections were created for Toad 10, Adobe Reader & Acrobat 8.2.1 and Adobe Reader & Acrobat 9.3.1. This feature is used to monitor and manage the usage of limited seat licenses.

One hundred and fifteen new Microsoft operating system security patches that were released were applied to the over 970 desktop computers on the NLM network. In addition, security updates and patches from other software vendors are also applied shortly after being released by their publishers. These patches are deployed overnight to NLM desktop systems to avoid user interruption and minimize downtime. Patches are then validated for effective application.

Testing was successfully completed with over 90 applications for compatibility with Internet Explorer 8 (IE8). Migration to IE8 will commence in FY2011.

Since the 2003 Help Desk consolidation with NIH's IT Help Desk, NLM desktop and PC networking support requests are now channeled to the NIH IT Help Desk for initial ticket entry into the call tracking system. Over 12,000 NLM ticket requests for IT support were entered and tracked. NLM IT staff resolved 75 percent of the calls (9,193 tickets) with 8 percent of support calls being completed by NIH staff. In addition, NLM conducted an awareness campaign to encourage users to contact the NLM support staff for all their IT support needs. As a result of this campaign, tickets increased by 28% from the previous year.

Three courses were coordinated this year including "Office Communication Server (OCS)," "Cognos," and "Altiris".

OCCS staff continued to support the NLM Outreach Program through volunteering and providing technical support to over 30 outreach events.

Public Health

Health Services Research Projects in Progress (HSRProj)

There are currently 8,616 projects in the Ongoing and Completed HSRPROJ file, as well as 12,084 archived projects. The HSRProj Web site was updated with 1,473 new records. Additionally, 1,125 records with the Final Date between 1/2005 and 12/2005 were relocated to the archive file. Modifications were made for processing updates to the HSRProj search site. The new procedure allows most of the updating to be done "offline," which is a more efficient process and the site only has to be unavailable for less than a minute. Modifications were made to the database tables in order to increase the efficiency of searches.

PHPartners.org & HSR Info Central

The Public Health Partners Web site is a site which helps the public health workforce find and use information effectively to improve and protect the public's health. This is a joint project among US government agencies (e.g. the Centers for Disease Control and Prevention, Agency for Healthcare 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).

The changes implemented to HSR Info Central included modifications to the FAQ page, the About HSRIC on the home page and the additional of a new widget for CER Queries on the home page. Also, a statistical report was created detailing organizations and the number of links associated with a given organization for each application.

Health Indicators Webinar

A new site was created for the National Information Center on Health Services Research and Health Care Technology (NICHSR) Health Indicators webinars, which includes a home page and user evaluation forms.

Web Exhibit

A new design interface began for the Outreach Activity Reporting Form (OARF) in order to support Web Exhibit and Consumer Health. A new Virtual Machine (VM) based system was set up. A prototype was designed for a desktop interface, combining OARF, Web Exhibit, Outreach Projects Database (OPD) and Outreach Contract Reports (OCR). Requirements gathering and consolidation will also be completed for the OPD re-design.

Outreach and Customer Services

Against the Odds Web Exhibit & Web Interactive Activities

The main accomplishment this year was the addition of the exhibition audio tour, which included the addition of new pages and navigation for the Audio Tour. A minimal flash audio player was also created in order to play the audio tour on demand within the site pages and on the new audio tour download pages. The "Voices From The Field" podcast was retrofitted.

Exhibit Asset Manager

The Exhibit Asset Manager database was created 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. A new feature was added this

year which allows the Image Management module to give HMD full control of the images associated with each object. To better define relations between contacts and assets, unified tagging was developed. The ability to label both assets and contacts with Exhibition level tags were implemented which will allow for better organization and categorization for reporting purposes.

Images from the History of Medicine (IHM)

This year, several batches of records were loaded into the new IHM system (LUNA). The LUNA load and indexing procedures were fine tuned as well as modification of IHM MARC records for reloading into Voyager.

HMD Directory

The HMD Directory is an online version of the History of Medicine component of DIRLINE (Directory of Information Resources Online), a National Library of Medicine (NLM) database, which contains location and descriptive information about a wide variety of health and biomedical resources. The DIRLINE History of Medicine component aims to assist scholars and researchers in identifying useful medical history collections throughout the world. Accomplishments include:

- Creating new tables, unique index files, and several other items on the search page.
- Writing programs to allow printing of all approved collections in PDF format, and all contact information in MS Word format.
- Adding statistical reports for date range of collections total, and for state/countries.
- Adding captcha when users submit a collection to avoid from submitting erroneous data.
- Adding three diacritics into MEDLINE Diacritics.

Customer Service Support System (Siebel)

Multiple defect resolutions, enhancement releases, and system upgrades were delivered for Customer Service, Change Request, and Firewall Service Request Management applications. Key accomplishments included:

- Adopting CanIt e-mail software to replace the lanlinux e-mail server as a part of system and application upgrade.
- Coordinating with PubMed and ClinicalTrials.gov web teams to create a backup to handle any power outage in the data center e-mails as a business continuity solution.
- Cleaning up the Siebel production file systems in a gateway server to meet the record retention requirements.
- Renewing SSL certificate in QA/Prod Web servers for secure client/server transactions.
- Completing the Siebel upgrade (from 7.7.2.8 to 8.1.1.1) in production as a technology refresh, which fixed many issues including the Citrix user login problems.
- Supporting customer service to process 96,995 service requests, 61,483 inbound e-mail messages, 53,234 outbound e-mail messages, 570 firewall change requests, and 1,062 application change requests.

Accessibility and Section 508

Section 508 is an amendment to the Workforce Rehabilitation Act of 1973 (29 U.S.C. § 794(d)) that was signed into law in 1998. Under Section 508, Federal Agencies must "give employees and members of the public with disabilities access to information that is comparable to the access available to others without disabilities." To comply with HHS' requirement to achieve full compliance and remediation with Section 508 by May 31, 2013, OCCS' major accomplishments for FY2010 included:

- Developed a comprehensive NLM Section 508 Compliance and Remediation plan;
- Added Section 508 conformance to NLM's "Policy on Broadcasting Administrative E-mail Messages;" and
- Trained over 60 NLM staff on the incorporation of Section 508 into the Acquisitions solicitation process.

ADMINISTRATION

Todd D. Danielson
Executive Officer

Table 32. Financial Resources and Allocations, FY2010
(Dollars in Thousands)

Budget Allocation:

Program Area	**Amount**
Extramural Programs	$66,666
Intramural Programs	259,638
Computer & Communications Systems	(29,334)
Library Operations	(74,278)
Lister Hill National Center for Biomedical Communications	(50,284)
National Center for Biotechnology Information	(87,277)
Specialized Information Services	(18,465)
Research Management and Support	14,082
Total Appropriation	340,386
Plus: Reimbursements	34,326
Total Resources	**$374,712**

Personnel

In October 2009, **Todd Danielson** was appointed NLM's associate director for administrative management, and is now a member of the Senior Executive Service. As associate director for administrative management, Mr. Danielson will continue to direct NLM's Office of Administration and will have an expanded role in policy and program development, implementation, and evaluation, in collaboration with other NLM senior staff members. Since joining the Library as executive officer in 2005, Todd has been an able leader of NLM's Office of Administration and a trusted advisor to the NLM staff. Among other achievements, he has recruited outstanding people for important vacancies, managed the establishment and expansion of NLM's Cooperative Acquisitions Center, overseen major enhancements to NLM facilities and the acquisition of additional space on and off the NIH campus, and guided NLM through the implementation of many new administrative systems and requirements.

In October 2009, **Rob Cohen** was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having been a TAJ Technologies contractor since June 15, 2009 at NCBI. Mr. Cohen received his master of science and bachelor of science degrees from the University of Maryland, College Park (1999, 1991). His work focused on developing tools and techniques for operating robots along with humans in space. Prior to joining NCBI, Mr. Cohen spent almost ten years at Immersion Medical building systems to simulate medical procedures used at academic medical centers to train physicians. In 2006, Mr. Cohen completed his MBA at the University of Maryland with a focus on Entrepreneurship and Management. He has authored numerous papers on space operations and is the inventor of more than a dozen US patents.

In November 2009, **Alexander Astashyn** was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having been an MSD/Lockheed Martin contractor since 2006 at NCBI. Mr. Astashyn earned a master's degree in Computer Science from the Polytechnic Institute of New York University in the field of machine learning. Having experience with scientific computing and mining large datasets after being employed at the Department of Veterans Affairs from 2004, he joined MSD in April, 2006 as a contractor for the National Center for Biotechnology Information (NCBI) to support the Gene and RefSeq projects. He helped develop and maintain data flows for warehousing, integration, and quality control of gene-centric information. He is also involved in the development of components for the framework for annotating genome assemblies (GPipe). As a staff scientist, he will continue supporting Gene, RefSeq, and GPipe projects.

In November 2009, **Nataliya Sharopova, PhD**, was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having been a TAJ Technologies Inc contractor since 2008 at NCBI. Dr. Sharopova earned a doctorate in genetics from N.I. Vavilov Institute of General Genetics (Moscow, Russia) for work in cotton genetics and phylogeny. Over the course of her career, she has worked in the fields of plant genetics and molecular biology, bioinformatics, and data management. She joined TAJ Technologies in October 2008 as a contractor for NCBI to provide scientific data management and pipeline development support for NCBI's database of Genotypes and Phenotypes (dbGap). As a staff scientist, she will continue this work along with other projects.

In November 2009, **Vitaly Stakhovsky** was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having worked as a Nanostock Inc./TAJ Technologies contractor assigned to NCBI since 2008. Mr. Stakhovsky earned a master of science degree in electronics engineering from the Moscow Institute of Physics and Technology. Over the course of his career, he has worked as a software engineer in several companies in Russia and the United States. He possesses broad experience in software development for Windows, from kernel device drivers to user interfaces. At NCBI, Mr. Stakhovsky has been responsible for the development and maintenance of projects such as Task Manager and the Aspera authentication module. As a staff scientist, he will continue to work on these and other internal NCBI projects.

In November 2009, **Dharitri Misra, PhD**, joined the Lister Hill Center as a staff scientist. She earned her

PhD and Master's degrees in Physics from the University of Maryland. Dr. Misra held senior positions in a number of research and development projects at the National Aeronautics and Space Administration (NASA), developing tools to automate and streamline advanced analysis functions for scientists and mission operators. At NLM, her research is focused on applying state-of-the-art digital preservation techniques to preserve NLM's digital collections, with special emphasis on automated metadata extraction from documents using machine learning tools.

In November 2009, **Sergey Y. Koshelkov** was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having been as an MSD/TAJ contractor at NCBI since 2001. Mr. Koshelkov earned a master of science degree from Moscow Power Engineering Institute for work in parallel computing. He joined MSD in December 2001 and was assigned to the Information Engineering Branch of NCBI as a contractor working on the PubMed Central (PMC) electronic archive. As a staff scientist, Mr. Koshelkov will be responsible for further optimization of PMC data management, storage, and data retrieval. He will also be involved in more advanced research projects aimed at designing and implementing novel algorithms for advanced full-text articles processed and indexed at NCBI (including PMC, PMCi, Bookshelf, and NIHMS-PMC integration).

In November 2009, **Brandi L. Kattman** was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist. Mrs. Kattman earned a master of science degree from Howard University for work in human genetics and in genetic counseling. Over the course of her career, she has worked in various areas of medical genetics, including patient care, disease research and clinical diagnostics. Most recently, she was the director of cardiology genetic services at GeneDx, a local diagnostic laboratory. Mrs. Kattman brings her expertise in clinical genetics and genetic testing to the NCBI where she will be responsible for overseeing the content and representation of medical genetic information. As a staff scientist, Mrs. Kattman will be responsible for reviewing information provided by existing databases at NCBI such as the Online Mendelian Inheritance in Man and GeneTests, and developing and implementing new projects such as the Clinical Variation database. Also, she will work with staff and external groups to maintain accurate and timely information about the genetic aspects of human disease and the laboratories providing testing for them.

In November 2009, **Jeffrey Reznick, PhD**, joined the History of Medicine Division of Library Operations as Deputy Chief. Dr. Reznick received his bachelor's degree in history at the University of Rochester, and his master's and doctoral degrees in history from Emory University. Dr. Reznick came to the History of Medicine Division from the American Occupational Therapy Foundation, where he served as director of the Institute for the Study of Occupation and Health. Previously, he was senior curator of the National Museum of Health and Medicine. Dr. Reznick has over a dozen years of administrative experience in the academic and nonprofit sectors, and his record of historical research is as extensive as his administrative career. He is the author of two scholarly monographs, *Healing the Nation: Soldiers and the Culture of Caregiving in Britain during the Great War* and *John Galsworthy and Disabled Soldiers of the Great War*, both with Manchester University Press, and of numerous articles, reviews, and book chapters. Dr. Reznick joins the National Library of Medicine with a longstanding interest in the institution, as both a patron through his research and as a member of the Friends of the NLM board for over seven years.

In December 2009, **Craig Wallin** was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having been a TAJ/Ariadne Genomics contractor since 2005 at NCBI. He earned his bachelor of science degree in computer science from the University of Illinois at Urbana-Champaign. He worked in compiler development for IBM at the Santa Teresa Laboratory in San Jose, California, and in software application development in Bethesda, Maryland. He also worked in software application development for ADP in Rockville. He joined Ariadne Genomics in August, 2005 as a contractor for the National Center for Biotechnology Information (NCBI) in the Information Engineering Branch to develop software applications for a number of projects, including Entrez Gene, Consensus Coding Sequence (CCDS), Genome Pipeline, and Gene Sensor. As a staff scientist, he will continue this work along with other projects.

In December 2009, **Deirdre (Dee) A. Clarkin** accepted the position of head, Collection Access Section, and (CAS) Public Services Division of Library Operations. In her previous position, Ms. Clarkin served as head of the Onsite Unit, CAS, at NLM, managing all aspects of a busy circulation department. In that capacity, she was responsible for the closed stack operation and delivering approximately 700-800 stack requests per day to patrons in the main Reading Room. She also served as government contracting officer technical representative for a multi-year contract that provides staff to oversee the photocopy and printing equipment and patron registration services for the main Reading Room. Prior to coming to NLM in 2002, Ms. Clarkin was a senior reference librarian for the Federal Deposit Insurance Corporation in Washington DC, providing research services to that staff nationwide with a special emphasis on legal and legislative history information.

In December 2009, **Melvin E. Hurr (Gene)** was appointed head of the Systems Services Section with the Systems Technology Branch of the Office of Computer and Communications Systems (OCCS). Gene brings a wealth of knowledge to the position in the areas of strategic IT planning, technical training and development, program and project management and data center consolidations. Gene earned a bachelor of arts from the

MidAmerican Nazarene University in 1996, followed by a MBA in 1998, also from MidAmerican Nazarene University. He also served in the United States Air Force for six years, stationed at Tyndall AFB in Florida, where he worked as a manager, shift supervisor, lead technician and Air Transport Specialist. Prior to joining the NLM/OCCS management team, Gene held the position of senior manager of data center and desktop operations at Sprint, where he was responsible for the support of all major business applications for Sprint PCS. While at Sprint, Gene also built and developed production and support models that were instrumental in the launch of 4G products.

In January 2010, **James Case, DVM, PhD**, joined the MeSH section of the Division of Library Operations as a health programs specialist, with principal duties of representing NLM and advising NLM on issues relating to the adoption and further development of SNOMED CT. Dr. Case was instrumental in the development of the National Animal Health Laboratory Network, which involved using standard terminologies from LOINC and from SNOMED CT. He has participated in standards activities for many years, and is active in HL7, having served as a co-chair of the HL7 Public Health and Emergency Response Special Interest Group for four years. Dr. Case's previous position was as Professor of clinical diagnostic informatics at the University of California, Davis.

In January 2010 **Jesus Cabaña, PhD,** joined the Lister Hill Center as a postdoctoral fellow. Dr. Cabaña completed a doctoral degree in computer science at the University of Maryland, Baltimore County in 2009. At NLM, he is working on medical imaging and 3D acquisition devices. His research involves designing new algorithms, tools, and techniques to explore large collections of unstructured and temporal medical images. In particular, a longitudinal and multimodal image dataset from the Center for Infectious Disease will be used to design tools to quantify, model, and visualize the progression of specific diseases. His LHNCBC mentor is Dr. Terry Yoo in the OHPCC branch.

In January 2010, **Bob Falk** was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having worked at Immersion Medical since 1998. Immersion Medical was the leading developer of Virtual Reality Medical Simulation products including simulators for laparoscopy, endoscopy, endovascular, and IV procedures. During his tenure there, Bob served in multiple roles including software engineer, project manager, and director of software development. Bob's development work at Immersion Medical focused on the implementation in C++ of real-time physics based simulations using OpenGL on a Windows platform. Prior to working at Immersion Medical, Bob worked for three years as an MSD contractor at the Lister Hill Center. While there, Bob developed database applications and visualization software in conjunction with the Visible Human Project. Bob holds a BA in computer science from the University of Michigan. As a staff scientist, Bob will be working at NCBI on the Genome Workbench team.

In January 2010, **Hani Z. Girgis, PhD,** was appointed to a research fellow position at the National Center for Biotechnology Information (NCBI). Dr. Girgis earned a doctorate in computer science from the University at Buffalo, The State University of New York, for work in machine learning applications to protein structure prediction. He joined the Johns Hopkins University in 2009 as a postdoctoral Fellow where he applied machine learning to develop a computer aided diagnosis system for Crohn's disease. As a research fellow at the NCBI, Dr. Girgis will continue his work on machine learning applications to develop computational methods to predict distant acting regulatory elements in the human genome.

In January 2010, **Laritza Taft, MD**, joined the Lister Hill Center as a staff scientist. Dr. Taft received her MD degree from the Universidad del Rosario in Bogotá, Colombia and is completing her PhD in biomedical informatics at the University of Utah. She was Head of the Department of Obstetrics and Gynecology at the University Hospital de San Jose in Colombia. Dr. Taft developed a data dictionary to map ICD-9 codes to clinical diagnosis and protocols for data capture for clinical studies and clinical academic research. At LHNCBC, she's contributing to a number of projects including medical vocabulary and messaging standards, Newborn Screening Coding and Terminology, and the NLM Personal Health Record.

In February 2010, **Swapna Abhyankar, MD**, joined the Lister Hill Center as a Clinical Postdoctoral Fellow. Dr. Abhyanka is a pediatrician who worked in private practice for five years and continues to practice one day every other month at Primary Pediatrics in Silver Spring, MD. She received her MD degree from the University of Michigan and completed a pediatrics internship and residency at Children's National Medical Center in Washington, DC. At NLM, she will be working on the NLM Personal Health Record (PHR) with Dr. Clem McDonald as her mentor. She will help develop the medical content contained in the PHR, including deciding what information from the patient's health history should be included in their personal health record and how that data should be entered, as well as creating preventive health reminders. She will also be working on standardizing newborn screening results reporting for the Newborn Screening Coding and Terminology Project.

In February 2010, **Terry Ahmed** accepted the position of Head, Reference and Web Services (RWS), Public Services Division in Library Operations. In his previous position, Mr. Ahmed served as head of the User Services Unit, managing all aspects of a busy customer service desk and the contract activities that support that service. He also managed reference services in the Main Reading Room. Prior to coming to NLM in 2001, Mr. Ahmed was the project manager of the customer services contract for Betah Associates. In that role he recruited, hired, trained and evaluated the staff and assisted all staff

with Web site navigation, database searching and preparation of correspondence to NLM customers. Prior to that position, Mr. Ahmed worked for seven years at Otsuka America Pharmaceuticals, responsible for day-to-day operations of a research library that included document delivery, circulation, collection development, cataloging, with administrative oversight for the budget. He also worked for three years at Bristol-Myers Squibb in Buffalo, NY, responsible for managing the serials, circulation and document delivery for that Library. Mr. Ahmed received his master's degree in library science from the State University of New York at Buffalo and his undergraduate degree from Buffalo State College.

In January 2010, **Stefan Wuchty, PhD,** was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist position at NCBI. Dr. Wuchty earned a master's degree in biochemistry and a PhD in theoretical biochemistry from the University of Vienna, Austria and held research positions at the European Media Lab in Heidelberg, Germany, University of Notre Dame and Northwestern University. Prior to his engagement, he was a staff scientist at the National Cancer Institute (NCI), where he was responsible for cancer research, utilizing systems biology and bioinformatics approaches. In his current capacity at the Computational Biology Branch (CBB) of the NCBI, he will be taking the leadership in bridging CBB's systems biology research activities that involve molecular data from the longitudinal Framingham study.

In February 2010, **Marc Gwadz, PhD**, was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having been a ComputerCraft Corporation contractor since 2004 at NCBI. Dr. Gwadz earned a doctorate in molecular biology from Columbia University's Department of Microbiology in the Graduate School of Arts and Sciences. Following that, he pursued postdoctoral research in the Biotherapy Branch of the Laboratory of Molecular Biology in the National Cancer Institute in Bethesda, MD. His work in the laboratory investigated the structure/function relationships of proteins, and the design and production of chimeric proteins for therapeutic applications. In 2004, Dr. Gwadz joined the Conserved Domain Database (CDD), part of the Computational Biology Branch of the NCBI. He joined a team of scientists that curates structure-aided, multiple alignment models of protein domain families to annotate key structural features within and explore evolutionary relationships between these families, and related projects. As a staff scientist, Dr. Gwadz will continue his work as a CDD curator in this project.

In February 2010, **Ning Ma, PhD,** was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having been a TAJ contractor since 2008 at NCBI. Dr. Ma earned a doctorate in physics from the Indiana University for work in the Hall effects in quantum bilayer systems. He has researched in computational physics at the West Virginia University as a postdoctoral fellow, focusing on the understanding and optimization of the thermal, mechanical, and electrochemical properties for materials to be used in future power plants and fuel cells. He joined TAJ, Inc. in July 2008 as a contractor for NCBI, to provide support in design, development, and implementation of tools and algorithms for the storage, retrieval, and analysis of information about molecular biology and genetics for the BLAST group in the Information Engineering Branch. As a staff scientist, he will continue this work along with other duties.

In March 2010, **Xia Jing, MD, PhD,** joined the Lister Hill Center as a clinical postdoctoral fellow. She has an MD degree from China and received the PhD degree in health informatics from the University of Salford, UK, in 2009. At NLM, Dr. Jing will explore a theoretical framework for clinical research informatics (CRI). She will work on clinical research design and on the analysis, interpretation, and dissemination of the information generated from clinical research. Her goal is to draw an overview map for research in CRI. She will also explore the use of a clinical decision support tool (namely clinical alerts) embedded in an electronic health record system (the Clinical Research Information System at the NIH Clinical Center) to improve clinical research workflow processes (such as clinical data collection and adherence to protocol procedures). Dr. James Cimino acts as Dr. Jing's mentor at NLM.

In April 2010, **Dachuan Zhang, PhD**, was appointed to the National Center for Biotechnology Information (NCBI) staff as a staff scientist after having been an MSD/Lockheed Martin contractor since 2004 at NCBI. Dr. Zhang earned a doctorate in biochemistry in Texas A&M University. After his doctoral study, he had postdoctoral research at Purdue University, where he focused on determination of protein crystal structures. His research seeks to better understand the relationship between protein structure and functions. Over the course of his career, he has worked on development and maintenance of the Gene Fragment Index database for microarray data management and analysis as a bioinformatics scientist at Gene Logic Inc. He joined MSD in March 2004 as a contractor for the National Center for Biotechnology Information (NCBI), to develop and maintain Web site servers for protein conserved domain, protein interaction, and molecular modeling databases, initially for protein structure data analysis. As a staff scientist, he will continue these works along with developing and maintaining NCBI protein structure Web site servers and developing algorithms for protein structure data analysis.

In April 2010, **Antonio José Jimeno Yepes, PhD**, joined NLM as a postdoctoral fellow. Dr. Jimeno Yepes completed a doctoral degree in computer science at the University Jaume I, Castellón de la Plana, Spain in 2009. At NLM, he is working on word sense disambiguation to improve the annotation performed by MetaMap. His research involves designing new

algorithms and techniques to disambiguate annotated terms in large collections of unstructured documents. His LHNCBC mentor is Dr. Alan R. Aronson in the CGSB branch.

In June 2010, **Di Huang, PhD,** was appointed to the National Center for Biotechnology Information (NCBI) as a Research Fellow after having been a postdoctoral research fellow at the Research Institute Methodist Hospital since 2008. She earned a doctorate in electrical engineering from the City University of Hong Kong, where she worked on machine learning and data mining. Dr. Huang worked on different topics of bioinformatics and computation biology (i.e., genomic and metabolic biomarks identification, gene regulatory networks of nuclear receptors and their roles on the development of diabetes type II, and the flux analysis of the metabolic networks, etc.) after graduation. As a research fellow, Dr. Huang will use her knowledge in bioinformatic data to reveal the regulatory mechanisms.

In June 2010, **Deborah Ozga** joined NLM as the new head of the Index Section. Since 2005, Ms. Ozga has served as a biomedical librarian at the National Institutes of Health Library, serving as the leader for a variety of projects. During her tenure at the NIH Library, she served as the project manager for the National Institutes of Health Library institutional repository. This responsibility included working with the NLM digital repository effort to develop detailed project requirements and the investigation, evaluation and recommendation of repository software. Ms. Ozga also served as the Team Leader of the NIH Library Instruction Team, offering classes on PubMed and other biomedical research databases, bringing a strong background in how users, both beginning and advanced, use MEDLINE and PubMed. Beginning in July 2009, she began a special detail assignment, requested by the Library of Congress, serving as a consortia coordinator. In this capacity, Ms. Ozga has been leading the effort to research the establishment of purchasing consortia to provide federal libraries with cost-effective access to resources, and thus bringing experience in working with and communicating with publishers. From 2002-2007, Ms. Ozga also served as a lecturer in information access at the University of Maryland's graduate program in library and information science. This course covered topics that included information search and retrieval skills, basic indexing principles, information seeking theories, and current issues in scholarly communications, institutional repositories, digital preservation, and the information industry.

In July 2010, **Pavel Ivanov** was appointed to the National Center for Biotechnology Information (NCBI) as a staff scientist after having been a Taj Technologies contractor since 2007 at NCBI. Mr. Ivanov earned an advanced engineering degree in information systems in radio physics and telecommunications from Nizhny Novgorod State University. He is proficient in several programming and scripting languages and has nine years of professional programming experience. Mr. Ivanov made significant contributions to developing NCBI's C++ Toolkit. He was also responsible for the NetCache distributed-storage server, database connectivity driver and client libraries, multi-threading extensions, unit test framework, and various issues related to template and preprocessor meta-programming in C++. As a staff scientist, he will continue this work along with other projects.

In July 2010, **Iris. Z. Zhu, PhD**, was appointed to the National Center for Biotechnology Information (NCBI) as a staff scientist after having been a Special Volunteer since Nov. 2009. Dr. Zhu earned a doctorate in biochemistry from the Institute of Biochemistry and Cell Biology of Chinese Academy of Science. In 2004, she started working in the mathematical modeling of biological systems in Cornell Medical College, where she focused on the theoretical simulation of dynamics of excitable physiological systems in the context of genetic and pharmacological interventions. After joining NCBI, she has been working on the statistical/machine-learning algorithms for RNA-seq analysis. She's been working closely with IEB branch to develop software product capable of generating full-length transcripts from RNA-seq experimental data. As a staff scientist, she'll continue this work along with other projects.

In July 2010, **Payel Ghosh, PhD**, joined the National Library of Medicine as Postdoctoral Fellow. She earned her PhD degree in Electrical and Computer Engineering from Portland State University, Portland, OR in June, 2010. Her research interests include medical image processing, machine learning, pattern recognition, and evolutionary computation methods such as genetic algorithms. She is currently working on local region-of-interest based image retrieval, cross-modal and semantic image retrieval. Her mentor is Dr. Sameer Antani in the Communications Engineering branch of the Lister Hill National Center for Biomedical Communications.

In August 2010, **Vladimir G. Ivanovskiy** was appointed to the National Center for Biotechnology Information (NCBI) as a staff scientist after having been a TAJ Technologies, Inc. contractor since March 2010 at NCBI. Mr. Ivanovskiy earned a master of science degree from the Moscow Institute of Physics and Technology State University in computer science. He started his career as a software developer in the Institute for System Programming Russian Academy of Science and participated in scientific research in the Compiler Technologies Department and commercial projects as a Telelogic AB subcontractor. Since January 2002, he worked in (NCBI) as an MSD/Lockheed Martin contactor. He designed and developed taxonomy software, authentication and SOAP web services, distributed task management software. In April 2008, he joined PCTEL RF Solutions company, where he designed and developed RF Data Collection software and RF Interference Management software. As a staff scientist, he will continue database and backends software design and development for MyNCBI and Author ID projects in PubMed group.

In August 2010, **J. Rodney Brister, MS, PhD,** was appointed to the National Center for Biotechnology Information (NCBI) as a staff scientist after being a Computercraft contractor at NCBI since 2009. Dr. Brister earned a master's degree in biology from Emory University and a doctorate in molecular genetics and microbiology from the University of Florida. His graduate work included molecular and biochemical characterization of Adeno-associated virus replication. Dr. Brister did postdoctoral work at the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), where he studied the replication of Bacteriophage T4 and was the recipient of the Nancy Nossal Fellowship Award. He joined Computercraft in February 2009 as a viral genome curator for NCBI and has worked on the SARS-CoV, Virus Variation, and Protein Clusters resources. He was also the chair of the Virus Working Group at the recent NCBI Annotation Workshop. As a staff scientist, he will continue this work along with other projects.

NLM Associate Fellows Program for 2010-2011

The NLM Associate Fellowship Program is an annual post-master's program for recent graduates of library programs. Associate Fellows receive a comprehensive introduction to NLM programs and services through a structured five-month curriculum phase, followed by a seven-month individual project phase. Projects selected by the Associate Fellows range across the scope and breadth of NLM, from production to research and development. Six new Associate Fellows began their year at NLM on September 1, 2010.

Julie Adamo received her MLS from the University of North Carolina, Chapel Hill in 2010. While working on her degree, Ms. Adamo worked as a research assistant for UNC's GrantSource Library and as a knowledge services assistant for Ipas, a non-profit women's health research organization. She also completed internships in the Instruction and Outreach department at the Duke University Libraries and at UNC's Health Sciences Library. Prior to pursuing an MLS, she spent four years as an indexer and editor for the Alternative Press Index and had previously worked as a circulation Assistant for the Baltimore County Public Library. Ms. Adamo has a BA in literature from Antioch College.

Kristen Burgess received her MLS with a specialization in health informatics from the University of North Texas in 2010. She received her BA in international studies from Emory University, during which time she studied in France and interned at a primary school and health clinic in Uganda. While completing her MLS, Ms. Burgess worked as the senior administrator at KDH Research & Communication, a public health and public policy research institution. During this time she also interned with the MetaArchive Cooperative and the Morehouse School of Medicine Library in Atlanta, Georgia, and volunteered in the Family Library and the Inman Medical Library at Children's Healthcare of Atlanta at Egleston.

Kristen Greenland received her MLIS from the University of Washington in 2010. She received a PhD in molecular & cellular biology from the University of Washington in 2008, and a BA from Carleton College in 2001 with a major in biology. During her time in the MLIS program, she worked as a research scientist in the biochemistry department where she studied cell division regulation in yeast. Dr. Greenland has completed an internship at the National Network of Libraries of Medicine Pacific Northwest Regional Medical Library and participated in a study abroad program to the Netherlands in the summer of 2009.

Stephen Kiyoi received his MLIS from the University of California, Los Angeles with a focus on medical informatics, in 2010. While pursuing his degree, he worked as a public service assistant at UCLA's Louise M. Darling Biomedical Library and as a student assistant to Christine Borgman, Presidential Chair of the Information Studies department. He also interned at the Norris Medical Library at the University of Southern California. Mr. Kiyoi is a recipient of the Medical Library Group of Southern California and Arizona's student scholarship of 2010. He received his bachelor's degree from the California Polytechnic University, San Luis Obispo in English Literature.

Caitlin Sticco received her MLS in 2009, and a specialist certificate in library and information studies in 2010, both from the University of Wisconsin at Madison. She received her BS in biomedical science from Antioch College, and has worked as a technician in molecular and cellular biology. While completing her graduate education, Ms. Sticco worked for the Wisconsin Clearinghouse for Prevention Resources as assistant special librarian and the UW Laboratory for Optical and Computational Instrumentation as a bioimaging informatics research assistant. In 2010, she also received the C. Berger Group Entrepreneurial Promise Award, for reflecting an innovative spirit and showing promise for an outstanding career in a special library, non-traditional library setting, or as an entrepreneur.

Salima M'seffar is from Morocco and is an International Fellow. She received her BA in June 2004 from L'Ecole des Sciences de l'Information in Morocco with honors, and her certificate in Spanish in 2008. She joined the National Institute of Hygiene (Rabat, Morocco) where she is in charge of the library in November 2004. Since the National Institute of Hygiene started working on its transition to becoming the National Public Health Agency, Ms M'seffar has supported its decision making through online searching and survey, and supporting ISO quality certification in Laboratories.

Retirements and Separations

On December 31, 2009, **Lawrence C. Kingsland, III, PhD**, retired as the assistant director for applied

informatics at the National Library of Medicine (NLM) and the branch chief of the Computer Science Branch (CSB) in the Lister Hill National Center for Biomedical Communications. Dr. Kingsland received his master's and PhD in electrical engineering at the University of Missouri. He was a postdoctoral fellow in NLM's Training Program at the University of Missouri in their Information Science Group. Joining the Lister Hill Center as the chief of the Expert Systems Program in 1984, he became the branch chief of the CSB in 1987. Under his leadership, the CSB's projects encompassed medical artificial intelligence, drug interface terminology, and assisted Internet-based biomedical information retrieval, including the UMLS Metathesaurus and the NLM Gateway meta-search system, which currently offers users simultaneous search in 23 NLM information resources. He was one of the principals in the development of Internet Grateful Med, which received then-Vice President Al Gore's "Hammer" award. Dr. Kingsland has received numerous awards for his research in biomedical expert systems, including the NLM Director's Award, the NIH Merit Award, and the NIH Director's Award. He has served on the boards of directors of the American Medical Informatics Association (AMIA), the Symposium on Computer Applications in Medical Care (SCAMC), and the American Association for Medical Systems and Informatics (AAMSI). For more than 35 years, Dr. Kingsland has contributed to research in medical informatics, designing and implementing complex computer systems to provide information to the public.

On January 1, 2010, **James S. Main** retired as the branch chief of the Audiovisual Program Development Branch (APDB). Mr. Main earned a BS degree in electrical engineering from Worcester Polytechnic Institute. He served four years of active duty as a commissioned naval officer and worked as an electronics engineer with the Naval Electronics Security Engineering Center. He continued to serve as a Commander in the US Naval Reserve until 1992. He joined the Lister Hill Center in 1975 in the Communications Engineering Branch, where he designed and managed the installation of the broadcast studios for a series of satellite communications experiments, known as the Communications Technology Satellite (CTS) project. He then joined the Computer Technology Branch where he developed NLM's first experiments with optical disc storage. In January 1984, Mr. Main became the branch chief of the newly formed APDB. Under Mr. Main's leadership, APDB provided the expertise and support required for the communications and multimedia programs of the National Library of Medicine. Among his recent accomplishments was positioning the NLM as the first government agency to implement high definition video production and presentation capabilities. Mr. Main received the Regents Award for Scholarship or Technical Achievement for his creative leadership in developing biomedical electronic image systems and facilities, and has received many awards for specific projects and performance over his long career. His contributions to the research and development of visual communications will beneficially impact the success of NLM outreach and communications programs for many years.

On February 14, 2010, **Simon Liu, EdD, PhD**, was appointed the new Director of the National Agricultural Library (NAL) in Beltsville, MD. Formerly, Dr. Liu was the Director of Office of Computer and Communication Systems (OCCS), a position that he held since 2000. Dr. Liu earned multiple master's degrees in the areas of computer science, business management, and government from Indiana University, University of Maryland and Johns Hopkins, respectively. Dr. Liu also earned two doctoral degrees, an EdD in higher education and a PhD in computer science, both from George Washington University. Under the direction of Dr. Liu, OCCS was able to lead a number of information technology based projects, which increased the virtual security of NLM's most valued data. He proved to be a reliable and able leader through his project management, program development and executive leadership. Among other achievements, Dr. Liu was able to create a new Branch under OCCS, the Medical Language Branch, which focuses on electronic medical records and exploring virtual means to access medical information for research institutions, hospitals, private physicians and government agencies. As Director of NAL, Dr. Liu will lead the world's largest research library for agriculture and related subjects.

On September 1, 2010, **Elliot R. Siegel, PhD,** retired as associate director for health information Programs Development at the National Library of Medicine (NLM) and Director of the Office of Health Information Programs Development. He was appointed an NIH Scientific Executive in 2008. Dr. Siegel received an MA in industrial-social psychology and his PhD in communication research, both from Michigan State University. Dr. Siegel joined NLM in 1976 as senior information scientist in the Lister Hill National Center for Biomedical Communications. In 1982, Dr. Siegel moved to NLM's Office of the Director, serving first as special assistant for operations research, then as assistant director for planning and evaluation, until being appointed to the associate director position in 1992. In those roles, Dr. Siegel established the NLM-wide evaluation research program that supported the development of numerous information and communications technology innovations for both health professionals and the public. On behalf of the Board of Regents, Dr. Siegel put in place the long range planning function that for more than two decades has strategically set the goals and priorities for NLM's research and service activities. He played a major role key nationwide outreach initiatives for underrepresented minority and underserved populations, the HIV/AIDS community. Dr. Siegel is an elected fellow of the American Association for the Advancement of Science (AAAS) and the American College of Medical Informatics (ACMI).

Awards

The NLM Board of Regents Award for Scholarship or Technical Achievement is awarded to recognize and stimulate independent creativity leading to scholarly and/or technical achievements that enrich biomedicine. The recipients of the 2010 award were **Alejandro A. Schäffer, PhD,** for exceptional achievement in applying bioinformatics research tools to the understanding and treatment of human genetic diseases; and **Alan R. Aronson, PhD,** for the development of two natural language processing tools, MetaMap and the NLM Medical Test Indexer, which provide in-house assistance to NLM indexers.

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 2010 award was **Joyce E. Backus** for more than 10 years of technical oversight and programmatic direction of MedlinePlus, and for moving NLM's consumer health products into the national spotlight.

The NLM Director's Honor Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to **Gale A. Dutcher** and **Frederick B. Wood, DBA,** in recognition of their creative development efforts, sustained leadership, and significant accomplishments that have shaped and strengthened NLM's Outreach Programs for Native Americans over an extended period of time.

The NIH Merit Award was presented to three individuals. The individual recipients were: **Hua Florence Chang** for leading the effort to develop the Radiation Emergency Medical Management System, a unique resource designed to assist health professionals during a mass casualty; **Terry S. Luedtke** for organization-wide technical contributions and dedicated efforts supporting various NLM applications; and **Jerrard R. Sheehan** for superb contributions to the development and support of federal information policy that enhances access to scientific data and research results.

The NIH Director's Award was presented to two NLM staff as individual awards and three NLM staff as part of NIH group awards. Individual awards were presented to **George R. Thoma, PhD,** in recognition of his exemplary leadership and outstanding expertise in advancing the state of the art in biomedical imaging and communication; and **Deborah A. Zarin, MD,** in recognition of her inventiveness, vision, and continuous enthusiasm in leading ClinicalTrials.gov through a period of extraordinary growth in volume and scope. The National Human Genome Research Institute, NIH-GWAS Group Award included **Jayashri P. Mehta,** in recognition of outstanding efforts to compile and curate the NHGRI Genome-Wide Association Catalog for the benefit of the scientific community. The National Institute of Neurological Disorders and Stroke, NIH-Emergency Medical Group award included **Clement J. McDonald, MD,** in recognition of his success in engaging the academic community in identifying the opportunities and challenges of research in an emergency setting. The National Institute of Neurological Disorders and Stroke, NIH-VPEP Group Award included **Rebecca M. Goodwin, JD,** in recognition of her exceptional dedication to creating an opportunity for NIH non-professional/non-scientific staff to improve their communication skills through the Volunteer Program for English Proficiency.

The 2009 Senior Executive Service Presidential Rank Award, Meritorious Executive, was awarded to **Betsy L. Humphreys** for sustained superior accomplishment in the management of programs to the United States Government and for noteworthy achievement of quality and efficiency in the public service.

Table 33. FY2010 Full-Time Equivalents (Actual)

Program Area	Count
Office of the Director	9
Office of Health Information Programs Development	6
Office of Communications and Public Liaison	10
Office of Administration	61
Office of Computer and Communications Systems	51
Extramural Programs	15
Lister Hill National Center for Biomedical Communications	66
National Center for Biotechnology Information	256
Specialized Information Services	42
Library Operations	284
TOTAL FTEs	**800**

American Recovery and Reinvestment Act

On February 17, 2009, H.R.1 was signed into law as the American Recovery and Reinvestment Act (ARRA) in an unprecedented effort to jumpstart the economy, create or save millions of jobs, and put a down payment on addressing long-neglected challenges facing the nation. NIH directly received $10 billion of two-year funds, of which $7.4 billion was transferred directly to the ICs and Common Fund in proportion to FY2009 Appropriations. NLM's portion of this distribution was $83,643,000, of which $37,070,349 was obligated in FY2009. The balance of $46,572,651 was obligated in FY2010.

NLM Diversity Council

The Director's Employee Education Fund, coordinated by the NLM Diversity Council, has proven to be a successful continuing education mechanism for NLM employees. The fund was established in 1998 by the NLM Director Dr. Donald A.B. Lindberg to empower every employee with the ability to reach his/her fullest potential. In FY2010, the fund approved 56 staff to take 90 courses, the majority of which were undergraduate. The school with the largest number of NLM enrollees was the University of Maryland

(24), followed by Montgomery College (10), and Strayer University (6). Other institutions included: Catholic University, University of the District of Columbia, Bowie State, American University, George Mason University, FAES-NIH Campus, Carroll Community College, College of Notre Dame of Maryland, Coppin State University, Howard Community College, Walden University, Penn State, University of Maryland Baltimore County College and School of ADV International Studies. Course disciplines included: financial accounting, human anatomy and physiology, statistics for biomedical scientists, archival principles, practices and programs, nursing-health policy, and special topics in health-medical terminology, among others. Students enrolled in traditional classroom instruction as well as online courses. The Diversity Council continues its efforts to publicize the availability of the fund.

NLM 2010 Health Expo/Take Your Child to Work Event

On April 22nd 2010, the NLM Diversity Council sponsored the third annual "Healthy Lifestyles for You and Your Family Expo" in conjunction with NIH's "Take Your Child to Work Day." This was the third year for what has now become an annual event. Once again the NLM Diversity Council was joined by the NIH Diversity Council who sponsored a table and coloring event for the children attending the Expo, adding a diversity theme to the art work. The products of this event, which depicted children's ideas of how diversity looks to them, were saved for use in an upcoming NIH diversity event later in FY2011. The NLM Expo and events featured a full day's worth of educational, family-oriented activities focusing on healthy living styles for both adults and children, again flavored with a multi cultural theme. The Emmy Award-winning Food-Play Productions which uses both live theatre and interactive media to promote healthy eating and exercise activities were joined this year by the NLM sponsored events: Passport to Global Health; Totally Toxic; and Kids on the Grid. This year the Health Expo also featured Tae Kuk Martial Arts demonstrations, a traditional Chinese lion dance and a hip hop dance demonstration and classes, all promoting healthy exercise as a way of maintaining good health. An inflatable play area erected for the children to use during the event was a great attraction. This year's Expo, attended by more than 2,000 NIH employees and their families, is becoming one of the most popular events held at the NIH. With this overwhelming response, NLM planning a repeat performance next year.

Appendix 1: Regional Medical Libraries

1. **MIDDLE ATLANTIC REGION**
NYU Medical Center
423 East 23rd St
Floor 15 South
New York, NY 10010
Phone: (212) 263-2030 Fax: (212) 263-4258
States served: DE, NJ, NY, PA
URL: http://nnlm.gov/mar

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

3. **GREATER MIDWEST REGION**
University of Illinois at Chicago
Library of the Health Sciences (M/C 763)
1750 West Polk Street
Chicago, IL 60612-4330
(312) 996-2464 FAX (312) 996-2226
States served: IA, IL, IN, KY, MI, MN, ND, OH, SD, WI
URL: http://nnlm.gov/gmr

4. **MIDCONTINENTAL REGION**
University of Utah
Spencer S. Eccles Health Sciences Library
10 North 1900 East
Salt Lake City, Utah 84112-5890
Phone: (801) 587-3412
Fax: (801) 581-3632
States Served: CO, KS, MO, NE, UT, WY
URL: http://nnlm.gov/mcr

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

6. **PACIFIC NORTHWEST REGION**
University of Washington
Health Sciences Libraries and
Information Center
Box 357155
Seattle, WA 98195-7155
(206) 543-8262 FAX (206) 543-2469
States served: AK, ID, MT, OR, WA
URL: http://nnlm.gov/pnr

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

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

Appendix 2: Board of Regents

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

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

Appointed Members

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

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

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

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

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

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

ROSSITER, Louis F., PhD
Research Professor
The Thomas Jefferson Program in Public Policy
The College of William and Mary
Williamsburg, VA

STANLEY, Eileen H., MLS
Roseville, MN

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

EX OFFICIO MEMBERS

LIU, Simon Y., Ph.D.
Director
National Agricultural Library
US Department of Agriculture
Beltsville, MD

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

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

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

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

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

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

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

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

Appendix 3: 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.

Appointed Members

ASH, Joan S., PhD
Associate Director
Department of Medical Informatics and Clinical Epidemiology
Oregon Health & Sciences University
Portland, OR

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

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

HUFF, Stanley M., MD
Chief Medical Informatics Officer
Intermountain Health Care
Murray, UT

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

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

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

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

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

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

Appointed Members

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

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

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

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

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

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

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

Appendix 5: Biomedical Library and Informatics Review Committee

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

Members:

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

Appointed Members

ARONSKY, Dominik, MD, PhD
Associate Professor
Department of Biomedical Informatics
Eskind Biomedical Library
Vanderbilt University Medical Center
Nashville, TN

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

DUNKER, A. Keith, PhD
Professor, Informatics
Center for Computational Biology and Bioinformatics
Indiana University Schools of Medicine & Informatics
Indianapolis, IN

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

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

LIDDY, Elizabeth D., PhD
Dean and Trustee Professor
Center for Natural Language Processing
School of Information Studies
Syracuse University
Syracuse, NY

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

MANDL, Kenneth D., MD, MPH
Associate Professor
Children's Hospital Informatics Program
Children's Hospital Boston
Boston, MA

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

MENDONCA, Eneida A., MD, PhD
Associate Professor
Department of Pediatrics
Section of Hematology/Oncology
Computation Institute
University of Chicago
Chicago, IL

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

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

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

SHATKAY, Hagit, PhD
Associate Professor
Head, Computational Biology & Machine Learning Laboratory
School of Computing
Queen's University
Kingston, ON

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

SITTIG, Dean F., PhD
Associate Professor
The University of Texas
School of Health Information Sciences at Houston
UT Houston-Memorial Hermann Center for Healthcare and Quality
Houston, TX

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

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

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

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

Appendix 6: Literature Selection Technical Review Committee

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

FLEMING, David A., MD **(Chair)**
Professor and Chairman,
Department of Internal Medicine
Director, MU Center for Health Ethics
University of Missouri School of Medicine
Columbia, MO

Appointed Members

CHRISTOPHER, Mary M., PhD
Professor of Pathology
Dept. of Pathology, Microbiology & Immunology
School of Veterinary Medicine
University of California
Davis, CA

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

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

DU, Chunying, PhD
Associate Professor
Cancer and Cell Biology
University of Cincinnati
School of Medicine
Cincinnati, OH

ELPERN, David J., MD
Dermatologist
The Skin Clinic
Williamstown, MA

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

JACKSON, Gretchen P., MD, PhD
Assistant Professor
Surgery and Biomedical Informatics
Vanderbilt Children's Hospital
Nashville, TN

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

NORTON, Catherine N., MLS
Director, Information Technology
Marine Biological Laboratory
Woods Hole, MA

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

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

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

ZHANG, Ge, MD, PhD
Assistant Professor
Department of Biomedical Engineering
University of Akron
Akron, OH

Scientific Review Administrator

KOTZIN, Sheldon, MLS
Associate Director, Library Operations
National Library of Medicine, NIH
Bethesda, MD

Appendix 7: PubMed Central National Advisory Committee

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

WARD, Gary E., PhD **(Chair)**
Professor
Department of Microbiology & Molecular Genetics
University of Vermont
Burlington, VT

Appointed Members

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

ALIRE, Camila, EDD
Dean Emeritus
University Libraries
University of New Mexico &
Colorado State University
Sedalia, CO

BIRD, Christopher J., BA
Solicitor
Legal Department
Wellcome Trust
London, England
United Kingdom

BLANTON, Ronald E., MD
Professor of Medicine
Center for Global Health and Diseases
Case Western Reserve University
Cleveland, OH

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

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

KANN, Maricel G., PhD
Assistant Professor
Department of Biological Sciences
University of Maryland
Baltimore, MD

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

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

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

TANNER, R. Michael., PhD
Provost and Vice Chancellor for Academic Affairs
University of Illinois at Chicago
601 S. Morgan Street
Chicago, IL

VELTEROP, Johannes, PhD
Chief Executive Officer
Knewco, Inc.
Rockville, Maryland

WEINTRAUB, Susan T., PhD Professor,
Department of Biochemistry Director,
Mass Spectrometry Laboratory
The University of Texas Health Science Center
at San Antonio
San Antonio, TX

WILBANKS, John T., BA
Boston, MA

Executive Secretary

LIPMAN, David J., MD
Director
National Center for Biotechnology Information
National Library of Medicine
Bethesda, MD

Appendix 8: Organizational Acronyms and Initialisms Used in this Report

Acronym	Meaning of Acronym
AAHSL	Association of Academic Health Sciences Libraries
AAPA	American Academy of Physicians Assistants
ABC	Advanced Biomedical Tele-Collaboration Test Bed)
ACORN	Automatically Creating OLDMEDLINE Records for NLM
ACP	American College of Physicians
ACSI	American Customer Satisfaction Index
AFIP	Armed Forces Institute of Pathology
AG	Access Grid
AHIC	American Health Information Community
AHRQ	Agency for Healthcare Research and Quality
AIDSinfo	Acquired Immune Deficiency Syndrome*info* (database)
ALTBIB	Alternatives to Animal Testing
AME	Automated Metadata Extraction
AMIA	American Medical Informatics Association
AMPA	American Medical Publishers Association
AMWA	American Medical Women's Association
APDB	Audiovisual Program Development Branch
APIRE	American Psychiatric Institute for Research and Education
ARRA	American Recovery and Reinvestment Act
ASCCP	American Society for Cervical Pathology and Colposcopy
ASPR	Assistant Secretary for Preparedness and Response, HHS Office of the
BAC	Bacterial Artificial Chromosome
BarSTool	Barcode Submission Tool
BGMUT	Blood Group Antigen Gene Mutation Database
BHEPP	Bethesda Hospitals' Emergency Preparedness Partnership
BISTI	Biomedical Information Science and Technology Initiative
BITA	Biomedical Image Transmission via Advanced Networks
BLAST	Basic Local Alignment Search Tool
BLIRC	Biomedical Library and Informatics Review Committee
BMT	Boundary Marking Tool
BOR	Board of Regents
BSAT	BMT Study Administration Tool
BoSC	Board of Scientific Counselors
BSD	Bibliographic Services Division
BSN	Bioinformatics Support Network
CAM	Complementary and Alternative Medicine
C&A	Certification & Accreditation (audit)
CANDHI	Central American Network for Disaster and Health Information
CARe	Candidate Gene Association Resource project
CAS	Collection Access Section
CBB	Computational Biology Branch
CBIR	Content-Based Image Retrieval
CCB	Configuration Control Board
CCDS	Consensus CoDing Sequence
CCHIT	Commission for Healthcare Information Technology
CCR	Central Contractor Registration
CCRIS	Chemical Carcinogenesis Research Information System
CDD	Conserved Domain Database
cDNA	Complementary DNA
CEB	Communications Engineering Branch
CEL	Affymetrix Cell intensity (file)
CgSB	Cognitive Science Branch
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
CMS	Centers for Medicare and Medicaid Services
COOP	(NIH Pandemic Flu) Continuity of Operations Plan
CORE	Clinical Observations Recording and Encoding
CoreBio	Core Bioinformatics Facility
CPSC	Center for Public Service Communication
CPT	Current Procedural Terminology
CRAC	Computer Room Air Conditioner
CRI	Clinical Research Informatics
CRISP	Computer Retrieval of Information on Scientific Projects
CSB	Computer Science Branch
CSI	Commission on Systemic Interoperability
CSR	Center for Scientific Review
CT	Computer Tomography
CTS	Communications Technology Satellite

Acronym	Meaning of Acronym
CTSA	(NIH Roadmap) Clinical Translational Science Award Centers
CUIs	Concept Unique Identifiers
DAC	Data Access Committees
DAR	Data Access Request
DART/ETIC	Developmental and Reproductive Toxicology/Environmental Teratology Information
dbEST	Database of Expressed Sequence Tags Center
dbGaP	Database of Genotypes and Phenotypes
dbMHC	Database for the Major Histocompatibility Complex
dbRBC	Database of Red Blood Cells
dbSNP	Database of Single Nucleotide Polymorphism
DDBJ	DNA Data Bank of Japan
DDD	Drug Delivery Devices
DDoS	Distributed Denial of Service (attack)
DCMS	Data Creation and Maintenance System
DEAS	Division of Extramural Administrative Support
DHHS	Department of Health and Human Services
DICOM	Digital Imaging and Communications in Medicine
DIMRC	Disaster Information Management Research Center
DIRLINE	Directory of Information Resources Online
DLXS	Digital Library Extension Service
DNA	Deoxyribonucleic Acid
DPR	Digital Preservation Research
DRAGON	Dynamic Resource Allocation in GMPLS Optical Networks
DRESWG	Digital Repository Evaluation and Selection Working Group
DRIG	Digital Repository Implementation Group
DTD	Document Type Definition
DVTS	Digital Video Transport System
EBI	European Bioinformatics Institute
EBP	Evidence-Based Practice
ECHO	European Community Humanitarian Office
Educollab	Educational Collaborators
EEO	Equal Employment Opportunity
EFTS	Electronic Funds Transfer Service
HER	Electronic Health Record
EMBL	European Molecular Biology Laboratory
EMR	Electronic Medical Record EMS Emergency Medical Services
EnHIP	Environmental Health Information Partnership
EnHIOP	Environmental Health Information Outreach Program
EP	Extramural Programs
EPA	Environmental Protection Agency
eRA	Electronic Research Administration
ESI	Early Stage Investigators
EST	Expressed Sequence Tag
ETIC	Environmental Teratology Information Center
EUREKA	Exceptional, Unconventional Research Enabling Knowledge Acceleration
FAES	Foundation for Advanced Education in the Sciences
FDA	Food and Drug Admnistration
FDCC	Federal Desktop Core Configuration
FHA	Federal Health Architecture
FIC	Fogarty International Center
FNLM	Friends of the National Library of Medicine
FTE	Full Time Employee
FTP	File Transfer Protocol
GAIN	Genetic Association Information Network
Gbps	Gigabits per Second
GCMS	Global Citation Management System
GDS	GEO DataSet
GEO	Gene Expression Omnibus (database)
GENSAT	Gene Expression Nervous System Atlas
GENE-TOX	Genetic Toxicology
GHR	Genetics Home Reference
GIS	Geographic Information System
GO	Grand Opportunity grant
GMAC	Grants Management Advisory Committee
GPS	Global Position System
SPU	Graphics Processing Unit
GRC	Genome Reference Consortium
GRMS	Global Records Management System
GSA	General Services Administration
GSS	Genome Survey Sequences
GTR	Genetic Testing Registry
GUI	Graphic User Interface
GWAS	Genome Wide Association Studies
HapMap	Haplotype Map
HAVnet	Haptic Audio Video Network for Education Technology
HBCU	Historically Black Colleges and Universities
HD	High Definition
HHS	Health and Human Services
HIPAA	Health Insurance Portability and Accounting Act
HITSP	Healthcare Information Technology Standards Panel
HLA	Human Leukocyte Antigen
HL7	Health Leven Seven, Inc.
HMD	History of Medicine Division
HSDB	Hazardous Substances Data Bank
HPCC	High Performance Computing and Communications
HPV	Human Papillomavirus
HRSA	Health Resources and Services Administration

Acronym	Meaning of Acronym
HSRIC	HRS (Health Services Research) Information Central
HRSInfo	Health Services Research Information
HSRProj	Health Services Research Projects
HSRR	Health Services and Sciences Research Resources
HSTAT	Health Services and Technology Assessment Text
HTTP	Hypertext Transfer Protocol
HuGENet	Human Genome Epidemiology Network
I3	Image Indexing Initiative
IAIMS	Integrated Advanced Information Management Systems
IBIS	Inferred Biomolecular Interactions Server
ICD	International Classification of Diseases
ICMJE	International Committee of Medical Journal Editors
ICs	Institutes and Centers (of NIH)
ICT	Information and Communication Technologies
IDE	Integrated Development Environment
IDS	Intrusion Detection System
IE8	Internet Explorer 8
IEB	Information Engineering Branch
IGSTK	Image Guided Surgery Toolkit
IHTSDO	International Health Terminology Standards Development Organization
IHM	Images from the History of Medicine
ILL	Interlibrary Loan
ILS	Integrated Library System
IMPAC	Information Management Planning Analysis And Coordination
InCHIs	IUPAC International Identifiers
INDSC	International Nucleotide Sequence Database Collaboration (formerly DDBJ/EMBL/GenBank)
*info*SIDA	infoSíndrome de Inmunodeficiencia Adquirida (database)
IP	Interactive Publications
IPv6	Next Generation Internet, Version 6
IRB	Institutional Review Board
IRC	In-Row Coolers
IRIS	Integrated Risk Information System
IRMA	Image Retrieval for Medical Applications
ISO	International Organization for Standardization
ISTO	Image Storage and Transmission Optimization
IT	Information Technology
ITER	International Toxicity Estimates for Risk
ITK	Insight Toolkit
ITP	Informatics Training Program
IUPAC	International Union of Pure and Applied Chemistry
JDBC	Java Database Connectivity
JDI	Journal Descriptor Indexing
JDMS	Journal Descriptor Maintenance System
JRE	Java Runtime Environment
KEGG	Kyoto Encyclopedia of Genes and Genomes
KSS	Knowledge Source Server (data)
LactMed	Drugs and Lactation (database)
LAN	Local Area Network
LHNCBC	Lister Hill National Center for Biomedical Communications
LJI	List of Journals Indexed
LO	Library Operations
LOINC	Logical Observations Identifiers, Names, Codes
LPF	Lost Person Finder
LRP	Long Range Plan (NLM)
LSI	List of Serials Indexed
LSTRC	Literature Selection Technical Review Committee
LVG	Lexical Variant Generator
MARG	Medical Article Records Groundtruth
MARS	Medical Article Records System
MAX	Mid Atlantic Exchange, U. of Maryland
MDoT	MEDLINE Database on Tap
MDT	Multimedia Database Tool
MEDLARS	Medical Literature Analysis and Retrieval System
MEDLINE	MEDLARS Online
MegaBLAST	Basic Local Alignment Search Tool
MEME	Metathesaurus Editing and Maintenance Environment
MEO	Medical Education and Outreach
MeSH	Medical Subject Headings
MHC	Major Histocompatibility Complex
MID	Manuscript Identifiers
MICAD	Molecular Imaging and Contract Database
MIM	Mentoring In Medicine
MIM	Multilateral Initiative on Malaria
MIMCom	MIM Communications Working Group
MIN	Multiple Ingredient (term type), RxNorm
MIRS	Medical Information Retrieval System
MLA	Medical Library Association
MLAA	Medical Library Assistance Act
MLB	Medical Language Branch (database server)
MLP	Molecular Libraries Program (at NIH)
MMDB	Molecular Modeling DataBase
MMS	MEDLARS Management Section
MMTx	MetMap Technology Transfer
MOR	Medical Ontology Research
MOU	Memorandum of Understanding
MS	Mass Spectrometry
MTHSPL	Metathesaurus Structured Product Labels
MTI	Medical Text Indexer
MTMS	MeSH Translation Management System
NA-MIC	National Alliance of Medical Image Computing
NAS	National Academy of Sciences

Acronym	Meaning of Acronym
NCBC	National Center for Biomedical Computing
NCBI	National Center for Biotechnology Information
NCCS	NIHI Consolidated Collocation Site
NCHS	National Center for Health Statistics
NCMHD	National Center for Minority Health and Health Disparities
NCI	National Cancer Institute
NCRR	National Center for Research Resources
NCVHS	National Committee on Vital and Health Statistics
N*e*HC	National *e*-Health Collaborative
NEI	National Eye Institute
NGI	Next Generation Internet
NHANES	National Health and Nutrition Examination Surveys
NHGRI	National Human Genome Research Institute
NHIN	National Health Information Network
NHLBI	National Heart. Lung, and Blood Institute
NIA	National Institute on Aging
NIAID	National Institute of Allergy and Infectious Diseases
NIBIB	National Institute of Biomedical Imaging and Bioengineering
NICHD	National Institute of Child Health and Human Development
NICHSR	National Information Center on Health Services Research and Health Care Technology
NIDCD	National Institute on Deafness and other Communication Disorders
NIDCR	National Institute of Dental and Cranio-facial Research
NIDDK	National Institute of Diabetes, Digestive, and Kidney Diseases
NIEHS	National Institute of Environmental Health Sciences
NIGMS	National Institute of General Medical Sciences
NIH	National Institutes of Health
NIHMS	NIH Manuscript Submission
NIH PI	NIH Pathways to Independence Award
NIMH	National Institute of Mental Health
NINDS	National Institute of Neurological Disorders and Stoke
NIOSH	National Institute for Occupational Safety and Health
NIST	National Institute of Standards and Technology
NLM	National Library of Medicine
NLP	National Language Processing System
NN/LM	National Network of Libraries of Medicine
NNMC	National Naval Medical Center
NNO	National Network Office
NOSC	Network Operations and Security Center
NOVA	National Online Volumetric Archive
NRCBL	National Reference Center for Bioethics Literature
NSF	National Science Foundation
NTCC	National Online Training Center and Clearinghouse
OACF	Onsite Alternate Computing Facility
OAM	Office of Administrative Management
OARF	Outreach Activity Reporting System
OCCS	Office of Computers and Communications Systems
OCHD	Coordinating Committee on Outreach, Consumer Health and Health Disparities
OCPL	Office of Communication and Public Liaison
OCR	Optical Character Recognition
OD	Office of the Director
ODIMRC	Office of the Disaster Information Management Research Center
OERC	Outreach Evaluation Resource Center
ORF	Original Release Format
OHIPD	Office of Health Information Programs Development
OMD	Office of Management and Budget
OMIA	Online Inheritance in Animals (database)
OMIM	Online Mendelian Inheritance in Man (database)
OMSSA	Open Mass Spectrometry Search Algorithm
ONC	Office of National Coordinator (for Health Information Technology)
OPASI	Office of Portfolio Analysis and Strategic Initiatives
OPD	Outreach Products Database
OSA	Optical Society of America
ORWH	Office of Research on Women's Health
OSIRIS	Open Source Independent Review and Interpretation System
PAHO	Pan American Health Organization
PBM	Pharmacy Benefit Manager
PCA	Personal Computer Advisory Committee
PCR	Polymerase Chain Reaction
PDA	Personal Digital Assistant
PDR	Publisher Data Review
PDB	Protein Data Bank
PDF	Portable Document Format
PDL	Personal Digital Library
PHLIP	Public Health Law Information Project
PHII	Public Health Informatics Institute
PHP	Public Health Partners
PHR	Personal Health Record
PHS	Public Health Service
PI	Pathway to Independence award
PI	Principal Investigator
PICO	Patient/Population, Intervention, Comparison, and Outcome

Acronym	Meaning of Acronym
PID	Pathway Interaction Database (NCI)
PIN	Precise Ingredient (term type), RxNorm
PLAWAR*e*	Programmable Layered Architecture With Artistic Rendering
PMC	PubMed Central
PMCI	PubMed Central International
PMC ID	PubMed Central Identification (number)
PRS	Protocol Registration System
PSD	Public Services Division
PUG	PubChem Power User Gateway
QA	Quality Assurance
QCIM	Quarterly Cumulative Index Medicus
RAC	Real Application Clusters
RCDC	Research Condition and Disease Categorization
RCSB	Research Collaboratory for Structural Bioinformatics
RDMS	Rare Disease Maintenance System
RefSeq	Reference Sequence (database)
REMIM	Radiation Even Medical Management
RFA	Request for Applications
RFID	Radio Frequency Identification
RFP	Request for Proposals
RHIN	Refugee Health Information Network
RIDeM	Repository for Informed Decision Making
RML	Regional Medical Library
RNA	Ribonucleic Acid
RNAi	RNA Interference
RPS-BLAST	Reversed Position Specific BLAST
RQS	Request Submission and Tracking System
RRF	Rich Release Format
RSS	Really Simple Syndication
RTECS	Registry of Toxic Effects of Chemical Substances
RTLS	Real Time Location System
RWJF	Robert Wood Johnson Foundation
SAB	Source Abbreviations
SBIR	Small Business Innovation Research
SCR	(MeSH) Supplemental Chemical Records
SDK	Software Development Kit
SEF	Serials Extract File
SEIM	Security Event and Information Management System
SEP	Special Emphasis Panel
SIDA	Swedish International Development Agency
SIG	Special Interest Group
SII	Scalable Information Infrastructure
SIS	Specialized Information Services
SKR	Semantic Knowledge Representation
SMART	Scalable Medical Alert and Response Technology
SNOMEDCT	Systematized Nomenclature of Medicine Clinical Terms
SO	Signing Official
SOAP	Simple Object Oriented Protocol (formerly Simple Object Access Protocol)
SPER	System for the Preservation of Electronic Resources
SPIN	Shared Pathology Informatics Network
SPIRS	Spine Pathology Image Retrieval System
SRA	Short Read Archive
STB	Systems Technology Branch
STTP	Short-Term Trainee Program
STTR	Small Business Technology Transfer Research
STS	Sequence Tagged Site
SVM	Support Vector Machine
TBL	The bottom line
TDI	3D Informatics (Group)
TEHIP	Toxicology and Environmental Health Information Program
TERA	Toxicology Excellence for Risk Assment
TIE	Telemedicine Information Exchange
TIFF	Tagged Image File Format
TKMT	Traditional Korean Medical Terms
TILE	Text to Image Linking Engine
TIOP	Toxicology Information Outreach Project
TOXLINE	Toxicology Information Online
TOXNET	Toxicology Data Network
TPA	Third Party Annotation (database)
TREF	Terminology Representation and Exchange Format
TRI	The Toxics Release Inventory
TSA	Transcriptome Shotgun Assembly
TSD	Technical Services Division
TT	Teaching Tool
TTP	Turning the Pages
UID	Unique Identifier (PubMed)
UKPMC	United Kingdom PubMed Central
UMLS	Unified Medical Language System
UMLSKS	UMLS Knowledge Source Server
UN	United Nations
UNCFSP	United Negro College Fund Special Programs Corporation
UPS	Uninterrupted Power Supply
UTS	UMLS Terminal Server
UTS	UMLS Terminology Server
UTS	Unified Terminology Services
VAST	Vector Alignment Search Tool
VHP	Visible Human Project
VM	Virtual Machine
VM	Virtual Microscope

Acronym	Meaning of Acronym
VPN	Virtual Private Network
VS	Virtual Slides
VTS	Visual Triage Study (NCI)
WAI	WebMARS Assisted Indexing
WashCAS	Washington Area Computer Assisted Surgery
WebMARS	Web-based Medical Article Records System
WebMIRS	Web-based Medical Information Retrieval System
Web-STOC	Web-Services Technology Operations Center
WGS	Whole Genome Shotgun
WIISARD	Wireless Internet Information System for Medical Response in Disasters
WISER	Wireless Information System for Emergency Responders
WSD	Word Sense Disambiguation
WTEC	World Technology Evaluation Center
XML	eXtensible Markup Language
XSLT	eXtensible Stytlesheet Language Transformations
YEP	(MeSH) Year-end Processing