

NATIONAL INSTITUTES OF HEALTH

National Library of Medicine

Programs and Services

Fiscal Year 2011

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

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

# CONTENTS

<b>Preface</b> .....	<b>vi</b>
<b>Office of Health Information Programs Development</b> .....	<b>1</b>
Planning and Analysis .....	1
Outreach and Consumer Health .....	1
Evaluation .....	2
International Programs .....	3
MedlinePlus African Tutorials.....	4
African Medical Journal Editors Partnership Program .....	4
Trans-NIH Collaboration.....	5
<b>Library Operations</b> .....	<b>6</b>
Program Planning and Management.....	6
Collection Development and Management.....	7
Selection and Acquisitions .....	7
Preservation and Collection Management.....	8
Vocabulary Development and Standards.....	9
Bibliographic Control.....	11
Information Products .....	12
Direct User Services.....	14
Outreach and Training .....	15
<b>Specialized Information Services</b> .....	<b>24</b>
Toxicology and Environmental Health Resources .....	24
Outreach Initiatives .....	26
Special Population Web Sites .....	27
Social Media .....	27
K-12 Initiatives .....	27
Environmental Health Student Portal.....	28
HIV/AIDS .....	28
Mobile Technologies .....	29
Disaster Information.....	29
Evaluation and Research.....	31
<b>Lister Hill National Center for Biomedical Communications</b> .....	<b>32</b>
Next Generation Electronic Health Records to Facilitate Patient-centric Care, Clinical Research, and Public Health .....	32
Biomedical Imaging, Multimedia, and 3D Imaging.....	33
Natural Language Processing and Text Mining.....	39
Information Resource Delivery for Researchers, Care Providers, and the Public.....	43
Clinical Vocabulary Standards and Associated Tools.....	45
RxNav.....	45
Communication Infrastructure Research and Tools.....	47
Video Production, Retrieval, and Reuse Project.....	49
Digital Video Archive .....	50
Training and Education at LHCNBC .....	50
<b>National Center for Biotechnology Information</b> .....	<b>51</b>
Molecular Biology Information Resources.....	51
Genome Information Resources.....	52
Integration of Clinical, Genetic, and Environmental Databases.....	59
Entrez Retrieval System .....	59
Discovery Initiative .....	60
Literature Information Resources .....	60
Research.....	61
Bioinformatics Training and Support.....	62

<b>Extramural Programs.....</b>	<b>64</b>
Overview of FY2011 .....	64
Success Rates of Grant Applicants.....	65
Research Grants .....	65
Resource Grants .....	66
Training and Career Awards .....	66
Career Support .....	67
Pan-NIH Projects .....	69
Extramural Programs Web Site.....	69
EP Operating Units.....	69
<b>Office of Computer and Communications Systems.....</b>	<b>79</b>
Consumer Health and Public Health .....	79
Controlled Medical Vocabularies .....	81
IT Infrastructure Services .....	82
Medical Literature Support and Document Delivery Services .....	85
Outreach and Customer Services .....	86
<b>Administration.....</b>	<b>88</b>
Personnel.....	88
NLM Diversity Council.....	93
<b>NLM Organization Chart .....</b>	<b>(Inside back cover)</b>
<b>Appendixes</b>	
1. Regional Medical Libraries .....	94
2. Board of Regents .....	95
3. Board of Scientific Counselors, Lister Hill Center for Biomedical Communications.....	97
4. Board of Scientific Counselors, National Center for Biotechnology Information.....	98
5. Biomedical Library and Informatics Review Committee.....	99
6. Literature Selection Technical Review Committee .....	101
7. PubMed Central National Advisory Committee.....	102
8. Organizational Acronyms and Initialisms   Used in this Report .....	103
<b>Figures</b>	
Figure 1. NLM Grants Budget - FY2007-FY2011 .....	64
Figure 2. FY2011 NLM T15 Training Program.....	67
Figure 3. NLM Sponsored Trainees at 18 Institutions .....	68
Figure 4. FY2011 EP Grants budget .....	73
Figure 5. FY2011 Research Grant Investment Areas.....	73
Figure 6. FY2010 Research Grant Investment Areas.....	73
<b>Tables</b>	
Table 1. Collections.....	19
Table 2. Collection Activities.....	20
Table 3. Cataloging and Indexing .....	21
Table 4. Services to the Public .....	22
Table 5. Selected Web Resources .....	23
Table 6. Success Rates - Core NLM Grant Programs FY2009 through FY2011 .....	65
Table 7. Extramural Programs Grants Budget by NIH Mechanism Groupings, and by Activity Code-FY2011 ....	72
Table 8. RFA/PA Actions in FY2011 .....	74
Table 9. Top Journals in which NLM Grantees and Trainees Published in FY2011 .....	75
Table 10. Five Most Frequently Cited Articles Acknowledging NLM Funding in FY2011 ( <i>Scopus</i> ).....	75
Table 11. NLM New Grants Awarded in FY2011 .....	76
Table 12. Financial Resources and Allocations, FY2011 ( <i>Dollars in Thousands</i> ).....	88
Table 13. FY2011 Full-Time Equivalents (Actual).....	93

## PREFACE

Fiscal Year 2011 marked NLM's 175<sup>th</sup> anniversary. A new Web site ([www.nlm.nih.gov/175](http://www.nlm.nih.gov/175)) captured the many milestones in our history of public service and information innovation, profiled the Divisions and chronicled anniversary events. Notable achievements in Fiscal Year 2011 include the following:

- In November 2010, we launched MedlinePlus Connect, a free service that brings persons using Electronic Health Record systems or patient portals to targeted consumer health information. It was named a "Secretary's Pick" for the coveted HHSinnovates awards. In a later round, NLM's Video Search software, part of NLM's Digital Collections repository, was named a finalist.
- On January 2011, the Library released a new design and organization for its main Web site. The NLM home page ([www.nlm.nih.gov](http://www.nlm.nih.gov)) has an entirely new layout, focused on users' top tasks.
- Two 175th anniversary contests engaged users of NLM data from around the US. "Show Off Your Apps," our first software development challenge, received many creative entries, with the winners invited to a grand event here. The "NLM & You" challenge invited the public to create original videos about how they use NLM services. Entries were uploaded to NLM's YouTube channel, which was launched this year.
- PubMed now contains over 21 million citations from more than 36,772 journals. In March 2011, PubMed Mobile Beta was released, making PubMed faster to load and easier to use on mobile devices.
- With the advent of high-throughput sequencing technologies, NCBI saw the Sequence Read Archive (SRA) double, to over 100 trillion bases of DNA. SRA is NIH's primary archive of data from "next-generation" sequencing technologies, which have revolutionized genomic research by increasing the speed of sequencing while reducing its cost. In addition, the NCBI Gene database grew to over 8 million genes from over 8,400 organisms, providing a single source reference for authoritative, curated descriptions on gene location and function.
- Disaster information remained an important concern. We launched the CHEMM tool, as well as special resources in response to the earthquake, tsunami and radiation incident in Japan. The Library also held its first Disaster Information Outreach Symposium in March 2011. This well-attended session was designed for and about information professionals meeting disaster information needs.
- May 2011 saw the awarding of new five-year contracts to eight institutions to serve as Regional Medical Libraries (RMLs) in the National Network of Libraries of Medicine (NN/LM). The NN/LM consists of the eight competitively selected Regional Medical Libraries and over 6,000 health sciences and public libraries.
- NLM was among a small group of US government agencies that participated in World IPv6 Day, June 8, 2011. Internet Protocol version 6 (IPv6) is the newest version of IP, succeeding the 30-year-old IPv4. Enabled applications during the Library's successful 24-hour test included the main NLM Web site, MedlinePlus, Daily Med and the MeSH Browser.
- This year saw the 4,400-mile journey of a healing totem created for the October 2011 opening of a new exhibition, Native Voices: Native Peoples' Concepts of Health and Illness. This work of art will serve as an exhibition focal point, which will explore the interconnectedness of wellness, illness and cultural life for Native Americans, Alaska Natives, and Native Hawaiians. A blog documenting the journey was launched.

Even as we look back on 175 years of accomplishments, we at NLM are always looking forward to creating new and better programs and services. I believe that those visionaries in the Office of the Army Surgeon General, who began this enterprise in 1836, would praise the gifted staff, advisors and consultants of National Library of Medicine. Surely they'd express wonder and admiration for the depth and breadth of your achievements, as do I. Thank you.



Donald A.B. Lindberg, MD  
Director

# OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

*Michael F. Huerta, PhD*  
*Associate Director*

The Office of Health Information Programs Development (OHIPD) is responsible for 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
- 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.

In FY2011, OHIPD staff continued outreach initiatives intended to encourage underrepresented minority high school students to utilize NLM's health information resources and to pursue careers in medicine and the health sciences, carried out in collaboration with other divisions of NLM.

An 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," 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 health care professionals and helps them execute a plan for success, encourages and promotes sustainable interest and participation in health careers for underrepresented minority African-American and Hispanic students located at schools in New York City's Harlem and South Bronx. An additional MIM Program initiated in FY2009, Science and Health Career Exploration, continues to be funded by NLM and co-sponsored by the Friends of the NLM. This initiative reaches six public and charter schools with an after school program to enrich the high school biology curriculum and encourage enrollment in higher education programs leading to degrees in medicine, allied health professions, and medical librarianship. Principals, science teachers and guidance counselors from participating schools oversee 40 sessions of biology instruction in 12 organ systems, taught by visiting health professionals/mentors over a two-year period. The participating schools constitute an ideal laboratory in which to experiment with new approaches for stimulating and sustaining student interest in health careers, within the larger context of Science, Technology, Engineering and Mathematics (STEM) education in the US. During its three-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 2011 semester, the evaluation findings revealed continuous impressive gains in health care knowledge. Through a combination of personal intervention and online education resources, the program employs an innovative educational curriculum that also strengthens high school students' readiness to pursue health careers.

NLM also continues its focused efforts to meet the health information needs of the Hispanic population in Texas and elsewhere. In FY2011, NLM continued its support for the South Texas High School for the Health Professions, known as MedHigh, a magnet health high school in the Lower Rio Grand Valley of Texas. The MedHigh VIVA! Peer Tutors Program is an award winning effort to involve high school students in teaching their peers about online health information. The peer tutors also conduct outreach to the local community and sponsor annual online virtual conferences open to interested faculty, librarians, and students from high schools around the country. MedlinePlus en español is being emphasized where applicable. The project also includes a health careers tutoring component that involves the students, teachers, and guidance counselors. In FY2011, plans were made to launch a health literacy video challenge to encourage students in grades six-12 to develop videos promoting use of MedlinePlus.gov. Peer tutoring has been extended to other magnet high schools in the Lower Rio Grande Valley, where it continues to be an effective outreach program. A publication addressing the long-term outcomes of the project was published in 2011. (Olney CA, Hansen L, Vickman A, Reibman S, Wood FB, Siegel E. *J Med Libr Assoc.* 2011 Oct;99(4):317-20.)

#### *Native American Outreach*

In FY2011 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 10 pow-wows, mostly in the Mid-Atlantic area. An estimated 23,000 persons visited the NLM booth over the course of these pow-wows. These activities proved to be another viable way to bring NLM's health information to the attention of segments of the Native American community and the general public.

OHIPD also supported two projects in the Dakotas that resulted largely from the Native American Listening Circles conducted in prior years and have continued into FY2011:

At Cankdeska Cikana Community College (via the Greater Midwest RML), Spirit Lake Nation, Ft. Totten, ND, a continuing project supports improvements at the tribal college library and development of health information-related educational and community outreach activities run by the library;

At MHA Systems Inc., a tribal enterprise of the MHA Nation, a continuing economic development

outreach project provided outreach assistance to a tribal information technology company to promote 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. Projects in FY2011 included 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 NLM *Native Voices* exhibition. For the AIDS Abstracts Conversion Project, a total of approximately 30,000 abstracts have been completed since project inception.

#### *Native Voices Exhibition*

OHIPD staff also contributed to the planning and development of the NLM exhibition, *Native Voices*.

In FY2011, OHIPD participated with other NLM divisions in coordinating and facilitating a consultation activity with Upper Plains tribal healers in Bismarck, ND, and the MHA Nation in New Town, ND. OHIPD arranged for the NLM Director to conduct interviews with tribal spiritual leaders, traditional healers, and community providers on topics relevant to the NLM "Native Voices" exhibition. This consultation also included a meeting with tribal college presidents in Bismarck. This was the sixth such consultation activity, with five others having been conducted in prior years in Anchorage, AK; Santa Fe, NM; Seattle, WA; Choctaw, MS; and the Hawaiian Islands.

OHIPD projects related to exhibition development included coordinating content development for the Healing Ways Galleries and project management for five exhibition components: Hokule'a scale model voyaging canoe, Herb Kane paintings, native ceremonial objects, native medicinal plants, and the Native Voices reference library. OHIPD staff also contributed to the planning, coordination, and implementation of the formal opening of the exhibition on October 5, 2011.

#### **Evaluation**

##### *Web Evaluation*

The Internet and World Wide Web play a dominant role in dissemination of NLM information services, and the Web environment in which NLM operates is rapidly changing and intensely competitive. These two factors continue to support the need for comprehensive and dynamic NLM Web planning and evaluation process. The Web evaluation priorities of the OHIPD include both quantitative and qualitative metrics of Web usage and measures of customer perception and use of NLM Web sites. During FY2011, the OHIPD continued to pursue an integrated approach intended to encourage exchange of information and learning within NLM, and to 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 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 FY2011, OHIPD continued to coordinate NLM's use of the online Web user survey known as the American Customer Satisfaction Index (ACSI). The ACSI provides ongoing user feedback to NLM's Web site managers.

OHIPD staff participated in FY2011 on the Trans-NIH Web Analytics Project Team, which was charged with conducting an NIH-wide survey and assessment of NIH Web analytics opportunities and challenges.

In FY2011, a paper addressing NLM's experience with adapting Web evaluation methodologies for the changing Web environment was published. (Keselman A, Cid V, Perry M, Steinberg C, Wood FB, and Siegel ER. Adapting Web 1.0 Evaluation Techniques for E-Government in Second Life. *Journal of Usability Studies* 6(4):204-225, August 2011).

### **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, African librarians are beginning to play a more substantial role in the project with medical journal editors, and the medical students are being trained by librarians in searching NLM databases.

NLM's initiatives in global health include demonstration projects that relate to 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, diarrhea, mental health, and Burkitt's lymphoma; 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.

A paper describing NLM's experiences with promoting health information access in developing countries was published in FY2011. (Royall J, Lyon B. *Afr Health Sci.* 2011 Sep;11(3):457-63).

### *Information and Communications Technology (ICT)*

NLM continued support in FY2011 for the implementation of an electronic health management information system (eHMIS - [www.ehmis.net](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 that connected a health center to the system and made medical information available at point of care.

### *Network of African Librarians*

NLM continues its commitment to utilizing and expanding the leadership of a growing network of African librarians who have received training as NLM Associate Fellows (<http://www.nlm.nih.gov/about/training/associate/africanetwork.html>) or as Medical Library Association Cunningham Fellows. The objective 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 eight librarians from Kenya, Zambia, Mozambique, Mali, Nigeria, Morocco, Uganda, and Zimbabwe.

The NLM Associates program and the significant role librarian's 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, and in collaboration with Kenyatta University, NLM convened a ground-breaking meeting in FY2009 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. Based on the enthusiasm generated at this meeting, the librarians collaborated online to produce course modules, and the course was launched at the Association for Health Information and Libraries in Africa (AHILA) 2010 international congress in Ougadougou, Burkina Faso.

In FY2011, related workshops were held by the NLM Network of African Medical Librarians and Deans at the Kenyatta University Conference Center in Nairobi, Kenya. As a result of their collaboration, the network members have created a training manual for use in Schools of Health Sciences in Africa called *Finding, Organizing, and Using Medical Information: A Training Manual for Students, Researchers and Health Workers in Africa*. Available in print and electronic formats available free of charge, the manual comprises seven modules: Information Sources, Searching Tools, Electronic Information Searching Techniques, Intellectual Property Rights, Management of Information, Evaluating Electronic Resources, and Scholarly Communications. The librarians have also made this content available on YouTube. After developing PowerPoint presentations and scripts for creating seven short videos for YouTube, the videos were shot in high definition by staff at Kenyatta University. Two of the modules were shot in English and in French, and a summary of the manual was created in Portuguese.

At their respective libraries, these librarians continue to train faculty and students and engage 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 that can be accessed online from anywhere. There is also interest in use of this training 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.

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

### **MedlinePlus African Tutorials**

This project is another effort by NLM to reach the consumer/end user, no matter where that user is located. MedlinePlus African tutorials focus on tropical disease issues in developing country contexts. The first two tutorials were on malaria ([www.nlm.nih.gov/medlineplus/africa](http://www.nlm.nih.gov/medlineplus/africa)) and diarrhea and were developed with the Faculty of Medicine at Makerere University in Uganda, field tested, and distributed in FY2008. As of FY2011, these tutorials have been expanded to include versions in six local languages.

Another tutorial 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 were created by children from around the region, many victims of the conflict. Following field testing in several settings— NGO, cultural village, family protection unit of police, Army, detention prison, college for teachers, schools—the final version of the tutorial is being modified to use simpler language to address the fact that some terms were too clinical and not understood by laypeople. When completed, a final version of the tutorial as well as a training manual for community use will be published.

A collaboration also continues on community education materials for Burkitt's lymphoma, working with a principal investigator from the National Cancer Institute, an Ugandan team of medical staff at two hospitals, artists, and a medical student team. Completed materials include a 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*

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

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

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

### **African Medical Journal Editors Partnership Program**

This medical journal capacity building program ([www.ajpp-online.org](http://www.ajpp-online.org)) began by focusing on journals associated with MIM sites in Uganda, Ghana, Mali, and Malawi. Currently, it comprises editors of *Mali Medical Journal*, *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 that they can be 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.

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. Four of the journals are now indexed in MEDLINE.

#### **Trans-NIH Collaboration**

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

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

#### *Visitors*

In FY2011, the Office of Communications and Public Liaison and the History of Medicine Division's Exhibition Program arranged tours and special programs for visitors from the following 78 countries:

Afghanistan, Australia, Bangladesh, Belgium, Benin, Botswana, Brazil, Bulgaria, Burma, Burundi, Canada, Chile, China, Colombia, Cote d'Ivoire, Croatia, Czech Republic, Democratic Republic of Congo, Djibouti, Ecuador, Egypt, Eritrea, France, Gambia, Germany, Ghana, Haiti, Hungary, India, Indonesia, Iraq, Israel, Italy, Jamaica, Japan, Kenya, Korea, Kosovo, Laos, Lebanon, Macao, Malawi, Mali, Mexico, Morocco, Mozambique, Nepal, Netherlands, Niger, Nigeria, Oman, Pakistan, Palestinian Territories, Philippines, Republic of South Africa, Russia, Rwanda, Saudi Arabia, Senegal, Serbia, Sierra Leone, St. Lucia, Sweden, Sudan, Swaziland, Taiwan, Tanzania, Thailand, Trinidad and Tobago, Tunisia, Turkey, Turkmenistan, Uganda, United Kingdom, United States, Ukraine, Zimbabwe.

# LIBRARY OPERATIONS

*Sheldon Kotzin*  
*Associate Director*

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

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

LO employs a multidisciplinary staff of librarians, technical information specialists, subject experts, health professionals, historians, museum professionals, and technical and administrative support personnel and relies on the services of contractors from a broad skill base. 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 Office (NNO), and the National Information Center on Health Services Research and Health Care Technology (NICHSR); and a small administrative office of the Associate Director (ADLO). LO staff members participate actively in efforts to improve the quality of work life at NLM, including the Diversity Council and technical tools such as the wiki, SharePoint and NLM Intranet.

Most LO activities are critically dependent on automated systems developed and maintained by NLM's Office of Computer and Communications Systems (OCCS), National Center for Biotechnology Information

(NCBI), or Lister Hill National Center for Biomedical Communications (LHNCBC). LO staff work closely with these program areas on the design, development, and testing of new system features.

LO also participates with national information standards organizations in the development of standards related to preservation, bibliographic control, collection holdings, vocabulary control, and data exchange. In FY2011 more than fifteen standards actions were reviewed.

## **Program Planning and Management**

Priorities for LO programs are based upon the goals and objectives identified in the NLM Long Range Plan 2006-2016 and, where appropriate, in the NLM Strategic Plan to Reduce Racial and Ethnic Disparities. The current NLM Long Range Plan emphasizes making the results of research readily available from scientific data to published literature to patient and consumer health information. In FY2011, LO working groups made progress towards goals published in the LO Strategic Plan for 2010-2015. The three strategic directions in the plan are: Transforming Access to the Collection; Redesigning Systems and Workflows; and Developing a 21<sup>st</sup> Century Workforce.

In addressing the strategic direction of Transforming Access to the Collection, LO engaged R2, a library consulting firm, to study workflows and processes for acquisitions and cataloging and to make recommendations for improved efficiencies based on NLM needs and industry standard practices. After interviewing over 100 NLM staff and consulting with many LO managers, R2 Consultants delivered over 100 recommendations to NLM management in April and gave a presentation to all interviewees and LO managers in June of 2011. Overall, R2 Consultants concluded that, while NLM has many similarities to other major research libraries, it also plays unique roles as a national library, such as being the organization responsible for major biomedical resources and the host of the NIH public access repository. The consultant recommendations followed the principles of: building on NLM strengths, leveraging the legacy, defining NLM's digital collection, establishing limits, and developing Library Operations systems-wide thinking to improve efficiency. LO established an oversight group that identified high priorities from the recommendations and is advising and holding staff accountable for implementing the recommendations.

For the strategic direction of Redesigning Systems and Workflows, LO continued to review and revise policies, procedures, and services to reflect shifting workloads; to use electronic information to enhance basic operations and services; and to work with other NLM program areas to ensure permanent access to electronic information. In FY2011, LO's Administrative Office continued to assist managers, supervisors and staff with the transition to a range of new administrative systems and the challenges of changes in the functioning of human

resources within NIH. Though many of its efforts are directed toward creating and promoting use of electronic information resources and supporting NLM's high priority outreach initiatives, LO also devotes substantial resources and attention to the care and handling of extensive collections of physical library materials and to the space and environment in which staff and patrons work and collections are stored. LO also focused considerable attention on working with other NLM program areas to meet the expanded responsibility for distribution of standard clinical vocabularies within the UMLS Metathesaurus. In FY2011, LO also continued to make improvements to conditions in the existing NLM library building and to develop strategies for handling the projected growth of the collections until a new facility becomes available. These improvements and strategies are discussed elsewhere in this report.

LO began holding quarterly all staff meetings to further communication and awareness within the Division. LO continued to encourage its staff to take advantage of flexiplace work arrangements as appropriate. One hundred and thirty-nine LO employees work at home at least one day per week and another 40 have ad hoc telework agreements in place. LO also developed a succession plan for key positions within the division.

### Collection Development and Management

The NLM comprehensive collection of biomedical literature is the foundation for many of the Library's services. LO ensures that this collection meets the needs of current and future users by updating NLM's literature selection policy; acquiring and processing relevant literature in all languages and formats; organizing and maintaining the collection to facilitate current use; and preserving it for subsequent generations. At the end of FY2011, the NLM collection contained 2,698,603 volumes and 16,422,701 other physical items, including manuscripts, microforms, prints, photographs, audiovisuals, and electronic media.

### Selection

In FY2011, selectors worked on a number of projects to enhance the breadth and depth of the NLM collection. For example, staff undertook an extensive review of Chinese monographs and journals and identified many new titles for the collection reflecting the expansion of biomedical publishing from China. Several gift collections were processed that added many unique titles not widely held in the US. Gifts were received from the Pan American Health Organization Headquarters Library, Smithsonian Institution Library, the Naval Health Research Center, the Agency for Healthcare Research and Quality, the Environmental Protection Agency libraries, and NASA as well as other institutional and individual donors. In addition, the Public Health Services Historian's Office transferred the entire library of over 5,000 items and

historical artifacts to NLM. While most of the printed material was already in NLM's collections, LO added over 600 new monographic titles after review. Significant grey literature content in the areas of disaster management and emergency medicine, health policy, health technology assessment, health statistics, and global health issues were added following an extensive review of Internet sites and e-mail communication with US and international agencies.

### Acquisitions

TSD received and processed 137,730 contemporary physical items (books, serial issues, audiovisuals, and electronic media). The number of electronic-only serials continues to increase. The total number of electronic-only serials grew to over 2,000 by the end of FY2011 and electronic-only serials now represent approximately 10% of all currently acquired serials. In FY2011, 5,640 licensed and 3,807 free electronic journals were available to NLM users. A net total of 35,445 volumes and 1,144,491 other items (including non-print media, manuscripts, and pictures acquired by HMD) were added to the NLM collection.

To address budgetary constraints, selectors reviewed serial titles for possible cancellation. They examined non-indexed titles with zero or low interlibrary loan requests from 2008-2011. The serial titles cancelled from the General Collection represented a savings of over \$150,000. Staff also identified and cut serial subscriptions from the print Reference Collection, the Staff Library, and NLM office copies. Total cancellations made in time for FY2012 resulted in approximately \$390,000 in savings.

### Acquisitions - History of Medicine

Expanding NLM's world-renowned historical collection that dates from the eleventh century to the present, HMD this year acquired a variety of significant printed books, manuscripts, modern archives, images, and historical films. These acquisitions included:

- Martin Luther's *Ob man fur dem sterben fliehen muge [Whether one may flee from a plague]* (Wittenberg: Hans Lufft, 1527), which focuses on the use of hospitals, the placement of cemeteries, and the duties of clergy and other public servants not to flee a city when epidemics such as the plague attack a town;
- *Regimen sanitatis Salernitanum* (Leipzig: Martin Landsberg, ca. 1486-89), an incunable edition of the famous and often reprinted thirteenth-century code of health presenting common sense advice about daily life including the best foods to eat, foods that should be avoided, and the importance of hand washing;
- Bartholomaeus Anglicus' *De proprietatibus rerum* (Lyon: Petrus Ungarus, 1482), another incunable, and the first important encyclopedia of the Middle Ages containing important chapters on theology,

philosophy, medicine, astronomy, botany, geography, and mineralogy; and

- William Turner's *The names of herbes in Greke, Latin, Englishe Duche & Frenche wyth the commune names that Herbaries and Apotecaries use* (London: J. Day & Wm. Seres, 1548).

Additions to the archival and manuscript collections included the papers of:

- Robert Chanock, MD, Robert H. Purcell, MD, Albert Kapikian, MD, and Brian Murphy, MD, all virologists with the National Institute of Allergy and Infectious Diseases;
- Margaret Pittman, PhD, Chief, Laboratory of Biological Products, Division of Biological Services, NIH, a pioneer in biological standardization; and
- The H1N1 Influenza Collection, documenting the response of the US Department of Health and Human Services to the 2009 flu pandemic.

Additions to the prints and photographs collections included:

- a stamp collection of roughly 700,000 items on medical topics, assembled by Adolf W. Schwartz, MD, DDS, PhD, and presented by his heirs;
- a collection of medical fine arts prints by noted printmaker May Lesser, contributed by Roche Pharmaceuticals; and
- a photographic album from circa 1916 showing American relief efforts in Paris.

Additions to the historical audiovisuals collections included some 45 personal films of epidemiologist Telford H. Work, MD, chronicling his extensive travels throughout the world and donated by his widow, Martine Jozan Work, MD

#### *Automation*

The OpenURL link resolver product, SFX, from ExLibris, was implemented in June 2011. SFX enables linking to the full text of articles from citation databases, and it also provides an A-Z list of electronic journals available for NLM staff and Reading Room patrons, greatly enhancing the value of subscription databases, such as those provided through EBSCOhost and Proquest. For on-site access to PubMed, SFX replaces LinkOut.

#### **Preservation and Collection Management**

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

conditions, and disaster prevention and response. LO distributes data about what NLM has preserved to avoid duplicate effort by other libraries. To assist other libraries facing disasters, LO created and released an Emergency Preparedness and Response Web site and mobile version that provide expert guidance to library and museum staff about conserving and repairing materials in numerous formats including books, photographs, and films that have been damaged during disasters. These tools were helpful to librarians in Japan in the wake of the earthquake/tsunami disaster. LO also works with other NLM program areas to conduct experiments with new preservation techniques as warranted and to promote the use of more permanent media and archival-friendly formats in new biomedical publications.

In FY2011, LO bound 18,997 volumes, repaired 2,304 items in NLM's onsite book repair and conservation laboratory, made preservation copies of 594 motion pictures and videos, and completed conservation treatment of 322 items specifically for digitization. One hundred and fifteen US libraries donated over 1,500 issues or volumes of serials to fill gaps in the NLM collection.

In FY2011, work continued on the multiyear project to improve building infrastructure in the stacks and expand existing collection space by strengthening the B-2 level floor, following earlier strengthening of the B-3 level ceiling, and installing compact shelving. NLM also completed the renovation of the southwest quadrant of B-2 and installed approximately 31,000 linear feet of compact shelving. The completed work reduced the height of 31 ranges of compact shelving in the southeast quadrant of B-3 to provide ceiling space to install a new sprinkler system which brings these collection spaces to current building codes for fire prevention.

Dwindling space for growth of the NLM collection requires resources to plan for additional space and to shift collections to available space. Over the course of 175 years, the NLM has repeatedly outgrown its facilities as its collection grows. This unparalleled resource of books, journals, and other materials contains much that does not exist elsewhere and as hospital and research libraries face increasing budget and space constraints, causing them to discard much of their own print collections, the NLM collection will assume even greater importance to the nation.

#### *MedPrint*

In September, NLM and the NN/LM announced the launch of MedPrint, a new national cooperative medical serials print retention program. This program is designed to ensure the preservation of and continued access to selected biomedical journals in print. The MedPrint program is open to all US libraries in DOCLINE.

Approximately 250 journals selected from Abridged Index Medicus® (AIM) and scanned PMC (PubMed Central) titles will serve as the first set of materials to preserve in print. MedPrint participants will

volunteer to retain one or more titles included on the list for up to 25 years, ending in 2036. The goal of the program is to get commitments for 12 copies of each title with widespread geographic distribution. NLM's copy will be the 13th.

The MedPrint Web site includes detailed information about the program, as well as links to the downloadable title list and the formal agreement document which is to be signed by the participating institution and NLM.

### *Digitization*

Beginning in FY2010, the Public Services and History of Medicine Divisions focused on digitization for the Medical Heritage Library Project. The Medical Heritage Library project is a cooperative venture to digitize historical materials from the collections of NLM, the Countway Library at Harvard, the Harvey Cushing/John Hay Whitney Library at Yale, the Augustus C. Long Health Sciences Library at Columbia University, and the New York Public Library, funded by a grant from the Alfred P. Sloan Foundation. Over a two-year period from January 2010 – January 2012, the participants will create digital copies of 30,000 books published before 1923 and make them publicly available on the Internet Archive. NLM is contributing 5,700 monographs with imprints from the United States, Latin America, the Caribbean and Canada dating from 1610 – 1865. In the early months of the project, PSD staff prepared the digitization lab space, acquired and installed the scanning equipment, and set up workflows to prepare material for scanning and performed quality control of the images and metadata outputs. Scanning began in September 2010 and staff scanned 4,200 volumes, containing 840,000 page images during FY2011. These digital files became part of the Medicine in the Americas collection in the NLM Digital Collection.

PubMed Central, a digital archive of medical and life sciences journal literature developed by the National Center for Biotechnology Information (NCBI), is the primary site for NLM to ensure permanent electronic access to journals. The PSD has worked closely with NCBI since 2004 to scan and add the backfiles of journals depositing current issues in the digital archive. In FY2011, one final scanned journal was released in its entirety, the *International Journal of Experimental Pathology*, bringing the backfile scanning project to a close. By the end of FY2011, more than 1,215,000 articles had been scanned and made available from over 90 journals. PubMed Central users retrieved articles in those journals more than 3,680,000 times.

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

The preservation of electronic information presents unique challenges that are not yet fully understood and therefore remain unsolved. NLM addresses these challenges by

using its own electronic services and publications as test-beds and works with other libraries and interested organizations to develop, test, and implement strategies and standards for ensuring permanent access to electronic information. LO works closely with other NLM program areas on activities related to the preservation of digital materials.

One area of experimentation in the preservation of digitized and born digital content is the collaborative NLM Digital Collections, the NLM repository to preserve and provide access to digitized and born digital content not included in PubMed Central. In FY2011, it was expanded to include over 1,406 books and 18 films from the History of Medicine collection. TSD worked with OCCS to enhance the infrastructure of Digital Collections to implement a distributed architecture that improves system redundancy and reliability. Digital Collections also provides an improved end-user experience with richer browse options. The Department of Health and Human Services (HHS) named the NLM Video Search software, part of the NLM Digital Collections repository, one of six winners in the HHS *Innovates* contest, giving the project an honorable mention award. The software, developed by OCCS in collaboration with LO, enables users to conduct accurate full-text searches of video transcripts from within a full-featured video player, and pinpoints exactly where the terms searched for appear in a film's timeline.

### **Vocabulary Development and Standards**

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

NLM is the central coordinating body for clinical terminology standards within the HHS. LO, in partnership with LHNBC 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 activities in this area have intensified, particularly in the areas of quality

and performance measurement, lab services, and newborn screening.

National Information Center on Health Services Research and Health Care Technology (NICHSR) routinely contributes to efforts in 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 representatives from MeSH and LHNBC. 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 work with AHRQ to support efforts to promote patient safety and adverse event reporting. In FY2011, 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.

In FY2011, NLM established an interagency agreement with the Office of the National Coordinator for Health Information Technology (ONC) to collaborate in terminology standards enhancement and related projects that support meaningful use of electronic health records; effective health information exchange for health care, public health, and research; and the goals of the Standards and Interoperability framework. The agreement established a vehicle for the NLM to be reimbursed for specific tasks related to: (1) ensuring that standard clinical terminologies (SNOMED CT and its US extension, LOINC, RxNorm) evolve efficiently to meet US-wide needs; (2) coordinating the development of standard clinical terminologies to avoid duplication and gaps; (3) distributing clinical vocabularies, subsets, and value sets, including within the UMLS Metathesaurus; (4) aligning standard clinical terminologies with clinical messaging standards and mapping them to HIPAA code sets; and (5) performing other related tasks that promote useful standardization of content in electronic health records.

### *Medical Subject Headings (MeSH)*

Library Operations opened FY2011 with a celebration of the 50<sup>th</sup> anniversary of the first print MeSH, issued as part of the 1960 *Index Medicus*. On November 18, 2010, Robert Braude, PhD, delivered a celebratory lecture on the occasion, "50 Years of Medical Subject Headings: Past, Present, and Future Impact on Biomedical Information." Dr. Braude, is a former NLM indexer who served as the director of the University of Nebraska, Leon S. McGoogan Library of Medicine and later, as the Frances and John Loeb Librarian and Assistant Dean for Information Resources at the Samuel J. Wood/C.V. Starr Biomedical

Information Center at the Joan and Sanford I. Weill Medical College of Cornell University.

The 2012 edition of MeSH contains 26,581 main headings, 83 subheadings or qualifiers, and more than 203,236 supplementary concept records for chemicals, other substances, and diseases. For the 2012 edition, the MeSH Section added 454 new descriptors, replaced 42 descriptors with more up-to-date terminology, and deleted 15 descriptors.

MeSH continued the work begun in 2010 of 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 MeSH descriptors. The remainder were introduced as Supplementary Concept Records (SCRs) in MeSH 2011. As with all SCRs, each of these rare disease SCR terms was mapped to ("Heading Mapped to") at least one current MeSH disease term to help searching and indexing. For example, Snyder Robinson Syndrome is a new disease SCR that is mapped to the MeSH descriptor Mental Retardation, X-Linked. Because rare diseases are defined as having a prevalence of fewer than 200,000 affected individuals in the United States, they traditionally received less attention and are sometimes called orphan diseases. Having these terms available for indexing purposes will enhance a more precise retrieval of the rare disease articles and contribute to their identification.

Two notable updates in FY2011 involved Algae and Sex Disorders. Taxonomically algae are polyphyletic, or derived from more than one evolutionary ancestry, and therefore the descriptor Algae no longer easily fits into the MeSH trees. For 2011 MeSH, the descriptor Algae was deleted and its children distributed among the appropriate eukaryotic trees. Sex Disorders also required changes based on the updated classification and new nomenclature recommendations put forth by the 2006 *International Intersex Consensus Conference*.

### *Clinical Vocabularies*

The MeSH Section and its contractors also produce RxNorm, a clinical drug vocabulary that provides standardized names for use in prescribing medications. RxNorm is designated as a US government-wide target clinical vocabulary standard by the Secretary of Health and Human Services. RxNorm provides a mechanism for connecting information from different commercial drug information services. In FY2011, brand names were added from First DataBank and Reuters Red Book.

DailyMed® provides access to over 20,000 structured product labels (SPL) from the Food and Drug Administration. DailyMed Mobile was released January 31, 2011. It has a simplified design for easy search, retrieval and display of product labels using any Web-enabled mobile device. DailyMed had more than 117 million page views in FY2011.

NLM supports the continued development and free distribution of LOINC® (Logical Observation Identifiers Names and Codes) by the Regenstrief Institute at Indiana University. In 2010, LOINC was identified as one of several required standards for use in Electronic Health Records (EHRs) to achieve meaningful use, as defined by the ARRA-HITECH provisions.

In FY2011, NLM also continued to support the US-wide license for the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), a comprehensive clinical terminology owned by the International Health Terminology Standards Development Organisation (IHTSDO). NLM, on behalf of HHS, is the US member of the IHTSDO responsible for distribution of SNOMED CT within the US and for representing US interests in its continued development. 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. FY2011 saw several important activities that will improve the use of SNOMED CT in the US and elsewhere. NLM launched the beta version of the US SNOMED CT Content Request System and published the first two releases of the US Extension to SNOMED CT.

FY2011 also saw the first transfers of content and related tooling from one major user of SNOMED CT to the IHTSDO for the potential benefit of all users. In September 2010, Kaiser Permanente, the IHTSDO, and HHS announced Kaiser Permanente's donation to the IHTSDO and the NLM of its Convergent Medical Terminology (CMT) content and related tooling, in order to facilitate clinician adoption of EHRs capable of generating comparable data for health care and public health. Where appropriate, the CMT content is being incorporated into the International Release of SNOMED CT or the US Extension to SNOMED CT.

NLM continues as the lead in the development of mappings from standard vocabularies to administrative code sets (e.g., SNOMED CT to ICD-9-CM), to support implementation of these standards in clinical and billing systems, with the goal of improving the 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 as close to their original sources, patients, health care providers, laboratories, diagnostic devices, etc., as possible. In FY2011, NLM participated in several mapping projects, both nationally and internationally, to achieve these goals. Most importantly, NLM began work on a map from SNOMED CT to ICD-10-CM to support reimbursement.

#### *UMLS Metathesaurus*

The content editing of the Unified Medical Language System (UMLS) is managed by the MeSH Section, using systems developed by LHNCBC and OCCS and maintained by OCCS. At the close of FY2011, the Metathesaurus contained more than 2.6 million concepts

and 8.6 million concept names from 161 source vocabularies in 19 different languages. The following new sources and updates were included in the UMLS: Consumer Health Vocabulary, Gene Ontology, HUGO Gene Nomenclature, International Classification for Nursing Practices, ICD-9-CM, Logical Observation Identifiers Names and Codes (LOINC), Medical Dictionary for Regulatory Activities Terminology (MedDRA), MEDCIN, Medical Subject Headings, NCBI Taxonomy, NCI Thesaurus, Online Mendelian Inheritance in Man, RxNORM, and SNOMED CT.

#### **Bibliographic Control**

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

#### *Cataloging*

LO catalogs the biomedical literature acquired by NLM to document what is available in the Library's collection and to identify high-quality, relevant medical resources on the Web. NLM also works with the Library of Congress (LC) to provide cataloging-in-publication (CIP) data for US medical books prior to their publication. Libraries worldwide use the cataloging and authority data created by NLM and doing so minimizes the cost and effort of creating cataloging data by these health science libraries. Cataloging is performed by the TSD Cataloging Section, staff in HMD, and contractors.

Additionally, the Cataloging Section is responsible for maintaining and updating of the *NLM Classification*, the maintenance and updating of the NLM Catalog Data Type Definition (DTD), transformation of Machine-Readable Cataloging (MARC) data into other metadata formats, and name authority control for selected NLM Web Services. The Cataloging Section released the 60<sup>th</sup> anniversary edition of the *NLM Classification* with the 2011 edition in online and PDF formats providing updates to the WL (Nervous System) and WM (Psychiatry) schedules.

FY2011 saw the completion of the US testing of *RDA (Resource Description and Access)*; the new cataloging rules intended to replace the *Anglo-American Cataloging Rules, 2nd ed. (AACR2)* and analysis of the data from that test. The three national libraries LC, NLM, and the National Agricultural Library (NAL) led the test which included 23 other partners from a variety of libraries, museums, and archives. NLM had three

catalogers from TSD and one from HMD participate in the test and, in addition to the 25 common records cataloged by all test participants, NLM catalogers created 385 original bibliographic records and 944 authority records using RDA.

Three LO staff served on the US RDA Test Coordinating Committee, which reviewed the massive amount of test data (23,366 bibliographic and authority records plus 8,509 surveys) and issued a public report in June 2011, approved by the managers of the three national libraries. This report concluded that RDA should be implemented no sooner than January 2013, if certain tasks and action items were completed by then. Some of these action items include: rewording RDA instructions in clear, unambiguous, plain English; having the RDA elements sets and vocabularies completed and published on the Web; improving the functionality of the online RDA Toolkit; integrating complete RDA record examples in MARC and other encoding schema into the Toolkit; and demonstrating credible progress towards a replacement for MARC. The Coordinating Committee is now monitoring the progress of the action items and will set a definite implementation date for RDA based on that progress. NLM ceased creating records using the new rules in December 2010 and will not begin use of RDA until the community agrees upon an implementation date.

In FY2011, the Cataloging Section cataloged a total of 19,191 books, serials, electronic resources, and audiovisuals, exceeding the previous year's total by 13%. These figures include 2,348 CIP records created from publisher electronic galleys and 2,101 titles cataloged for the HMD collection. With a full complement of catalogers on board, contract cataloging is decreasing and more work is being done in-house.

HMD cataloged or substantially upgraded catalog records for 4,008 early printed monographs, 842.8 linear feet of manuscripts, 992 historical audiovisuals, and 704 prints and photographs.

### *Indexing*

LO indexed 5,559 biomedical journals for the MEDLINE/PubMed database to assist users in identifying articles on specific biomedical topics. The indexing workload continues to rise, partly due to the selection of new journals for MEDLINE/PubMed, but primarily due to increases in the number of articles published in journals already being indexed. In FY2011, a combination of in-house staff, contractors, and cooperating US and international institutions indexed 724,831 articles, a 4% increase from last year. Previously indexed citations were updated to reflect 271 retractions, 7,870 published corrections, and 35,049 comments found in subsequently published notices or articles.

In FY2011, indexers created 80,297 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI Gene database.

Index Section switched another 301 journals to process from the electronic version rather than the print version, resulting in a grand total of 4,053 titles for FY2011 being handled from the online version, representing 73% of all currently indexed journals.

Keeping up with the increases in the number of articles and journals that are indexed for MEDLINE requires new approaches to indexing to improve effectiveness and efficiency of the indexing process. LO continues to work with the LHCBC and OCCS on the Indexing 2015 initiative to identify, test, and implement ways to reduce or eliminate tasks now performed by human indexers. In FY2011, significant progress was made with the introduction of the Medical Text Indexer First Line (MTIFL) pilot project using the Medical Text Indexer (MTI) technology as the first indexing step. After this step, senior indexers review, validate, and improve on the MTI suggested indexing. By the end of FY2011, 2,849 articles from 20 journals had been processed using MTIFL.

Indexers perform their work after the initial data entry of citations and abstracts by one of two means: electronic submission from publishers, the fastest and most economical method, or scanning and optical character recognition (OCR). In FY2011, 92% of the citations and abstracts were received electronically from publishers and the remaining 8% were scanned.

During this fiscal year, NLM implemented a new policy for indexing electronic-only journals. The new policy clarified NLM's preservation, access and licensing requirements for electronic-only journals indexed in MEDLINE. To implement this new version of the policy, LO is working with journals newly approved for MEDLINE inclusion to ensure their compliance with the policy. MEDLINE journals that have become electronic-only publications will need to comply with the policy in the future.

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 FY2011, the Committee reviewed 558 journals. They recommended a total of 109 journals, including 11 titles that were recommended provisionally as they are available only in electronic form and must meet the new electronic journal policy.

### **Information Products**

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

### *Databases*

LO manages the creation and maintenance of the content of MEDLINE/PubMed NLM's database of indexed

citations; LocatorPlus and the NLM Catalog; MedlinePlus and MedlinePlus en español, NLM's primary information resources for patients, families, and the general public; and a number of specialized databases, including several in the fields of health services research, public health, and history of medicine. DailyMed, a database of structured product labels (SPL) approved by the Food and Drug Administration, provides access to the prescribing information approved by the FDA. These databases are richly interlinked with each other and with other important NLM resources, including PubMed Central, other genome resources, ClinicalTrials.gov, Genetics Home Reference, as well as Specialized Information Services (SIS) toxicology, environmental health, and AIDS information services.

Use of MEDLINE/PubMed was approximately 1.83 billion searches, a 13% increase from last year. Page views in PubMed totaled 4.78 billion. MEDLINE/PubMed now includes more than 21 million citations. Over 48,000 citations from 1946 *Current List of Medical Literature* (CLML) were added to PubMed in FY2011.

BSD staff assisted NCBI with the implementation of a new PubMed Mobile product, introduced in March. LO staff also assisted in the merger of the NLM Catalog and the Journals Database on the Entrez system, providing a unified interface for users. The NLM Databases and Electronic Resources Web page was transformed into a searchable database and renamed the Databases, Resources & APIs in order to provide better discovery of all the databases and electronic resources provided by NLM.

In another effort to showcase the variety of data and systems produced by NLM, the NLM Challenge called Show Off Your Apps: Innovative Uses of NLM Information Challenge was launched in April. Forty-two applications were submitted. Five winners and five honorable mentions were selected by outside Federal Advisory Committee members and recognized at an awards ceremony on November 2, 2011.

Library Operations brings consumer health information to the nation through the authoritative, up to date information in MedlinePlus and MedlinePlus en español. The entire suite of MedlinePlus products, including the full and mobile sites, MedlinePlus Connect, the health topics eXtensible Markup Language (XML) Web Service and Twitter followers saw tremendous use in FY2011. Over 164 million visitors viewed over 728 million pages on the MedlinePlus full Web site. MedlinePlus Mobile received over 650,000 visitors viewing over 2,800,000 pages. At the end of FY2011, there were 270,000 subscribers with over 34 million subscriptions to customizable e-mail updates of new content on MedlinePlus. MedlinePlus remains in the top five of government sites based on the American Customer Satisfaction Index (ACSI). PSD and OCCS continued to expand and improve the basic content and features of MedlinePlus.

In November 2010, NLM launched MedlinePlus Connect, a free service that brings patients or providers in

an Electronic Health Record (EHR) system or patient portal to targeted consumer health information from MedlinePlus. MedlinePlus Connect allows EHR systems to use existing code sets to link to relevant information from MedlinePlus.gov. There is no registration required, and MedlinePlus Connect is available as either a Web Application or a Web service. Patients can link to information on diagnoses, drugs and lab results. MedlinePlus Connect can be built into a patient portal or EHR software and offered to clients as a pre-existing feature or as an option to "turn on." Hooking up MedlinePlus Connect is a quick, one-time change for any health IT system. MedlinePlus Connect was recognized as a "Secretary's Pick" for the HHS *Innovates* awards in March 2011 followed by an HHS issued press release touting the importance of the service.

In order to simplify its naming, the MedlinePlus Twitter account, medlineplus4you, changed to @medlineplus and reached over 21,000 followers at the end of FY2011. MedlinePlus grew by 44 English health topics and 55 Spanish health topics offering a total of 903 English health topics and 881 in Spanish. Increasing the number and variety of health topics on MedlinePlus allows Connect to respond to more diagnosis requests from patient portals or EHR software.

NIHSeniorHealth grew by two topics in FY2011 bringing the total to 50 topics. Topics added were *Long-Term Care* from the Administration on Aging (AOA), and *Anxiety Disorders*, contributed by the National Institute for Mental Health (NIMH).

Under the direction of NICHSR, NLM continues to expand and enhance its Web-based resources and databases for health services researchers and public health professionals. The number of serials on topics related to health services research and public health that are indexed in MEDLINE continued to increase during FY2011.

NICHSR collaborated with NCBI to maintain and expand health services research (HSR) and public health resources available on Bookshelf. Additions to Health Services and Technology Assessment Text (HSTAT) on the Bookshelf included new documents from the Agency for Healthcare Research and Quality (AHRQ), from its Effectiveness Program and Evidence Based Practice Centers. Also new to HSTAT in FY2011 were two reports from the Office of the Surgeon General: *Report of the Surgeon General on How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease*, and *The Surgeon General's Call to Action to Support Breastfeeding* and 11 new reports from the US Preventive Services Task Force (USPSTF).

NCBI continued to expand the set of clinical guidelines-related materials on Bookshelf and released a new site, PubMed Health, drawing from such sources as the German Institute for Quality and Efficiency in Health Care (IQWiG), the National Institute for Health and Clinical Excellence (NICE) and the Institute of Medicine. NICHSR worked closely with NCBI throughout the year

to improve the design, format and content of PubMed Health.

NICHSR worked with AcademyHealth and a team of librarians and subject matter experts to update three core library education modules, (1) Health Policy (2) Health Economics and (3) Health Outcomes. These modules recommend selected books, journals, Internet sites, bibliographic databases, and organizations.

In FY2011, NICHSR made improvements to the search functionality of the HSR Info Central site. As a result of these efforts, and increased promotion of the site to the HSR community, use of the site, measured by visits and visitors, increased more than 41% from the end of FY2010 to the end of FY2011.

NICHSR worked with AcademyHealth and the Sheps Center at the University of North Carolina, Chapel Hill to expand Health Services Research Projects in Progress (HSRProj) and with Information Packaging and the HSRR (Health Services and Sciences Research Resources) database to include PHSSR (Public Health Systems and Services Research) datasets. In FY2011, the HSRProj database grew to more than 9,000 active records representing almost 120 funders, both international and domestic. In FY2011, the HSRR database increased to over 1,100 records describing datasets, survey instruments or software, of which almost 20 percent are explicitly related to public health systems and services research.

NICHSR was involved in a number of activities related to examining how the NLM resources could support increasing interest in Comparative Effectiveness Research (CER). In one such effort, PubMed, HSRProj, and ClinicalTrials.gov are key sources for a Department of Health and Human Services CER inventory being developed by the Office of the Assistant Secretary for Planning and Evaluation. The inventory is being designed for use by the general public, practitioners, and policy-makers. In FY2011, NICHSR and ADLO staff served on behalf of NLM as key technical advisors to this project. In FY2011, use of the NICHSR-developed and tested Comparative Effectiveness Research search strategies increased more than 10-fold over FY2010.

In FY2011, NICHSR hosted a two-part Webinar series about Health Technology Assessment (HTA), HTA 101 and Comparative Effectiveness Research conducted by Clifford Goodman, of the Lewin Group, a health policy consulting firm. NICHSR staff also updated the popular online training primer HTA 101: An Introduction to Health Technology Assessment.

#### *Machine-Readable Data*

NLM licenses many of its electronic databases to other organizations to promote broad use of its authoritative bibliographic, vocabulary, and factual data. There is no charge for any NLM database, but recipients must abide by use conditions. Commercial companies, universities, and other organizations that obtain NLM data incorporate them into many different database and software products and

use them in a variety of research and development projects.

Demand for MEDLINE/PubMed data in XML format continues to increase with 688 MEDLINE licensees in FY2011, an increase of 11% from the previous year. Of the 736 total licensees of MEDLINE and other NLM databases, 539 report research use. A relatively small number of organizations lease NLM catalog records or one or more of the SIS toxicological or environmental health files in XML format. At the end of FY2011, there were 6,566 UMLS Metathesaurus licensees, an increase of 7% from last year.

#### **Direct User Services**

LO retrieves materials requested by onsite patrons from NLM's closed stacks and also provides interlibrary loan as a backup to document delivery services available from other libraries and information suppliers. In FY2011, PSD's Collection Access Section processed 363,910 requests for contemporary documents. HMD handled 5,537 requests for rare books, manuscripts, prints, photographs, and historical audiovisuals.

The number of onsite requests in the NLM Main Reading Room continued to decline primarily due to online access to more journals in the Reading Room and free access to journals in PubMed Central. Onsite users requested 119,209 contemporary items from the stacks, an 18% decrease from last year. In March, PSD discontinued the Overnight Photocopy Service due to low use and high overhead. This service was implemented in the 1990's to accommodate users with high volume who did not want to make their own copies. Users of the HMD Reading Room requested 4,824 items from historical and special collections, a 35% decrease from last year.

In FY2011, the Collection Access Section processed a total of 244,701 interlibrary loan requests (ILL), a 5% decline from the previous year, continuing the decade long trend of decreasing demand for ILL service. The one-day processing rate remained at 98% and NLM delivered almost all, 97%, of ILLs electronically. In March, NLM increased the ILL loan period for books to 60 days in response to customer requests, and internal need for reducing the number of overdue notices and customer service contacts.

A total of 2,818 libraries use DOCLINE, NLM's interlibrary loan request routing and referral system. NLM released DOCLINE 4.6 which completes the redesign of the request module and includes changes to Loansome Doc transfer and the resubmit feature for retired requests. Late in FY2011, NLM released DOCLINE 4.7 with features to support the new national print retention initiative, MedPrint.

DOCLINE users entered 1.6 million requests in FY2011, an 8% decrease from the previous year; 93% of the requests were filled. The NLM share of all DOCLINE requests remained steady at 11% of the total. Individuals submitted 296,836 requests to DOCLINE libraries via the

Loansome Doc feature in MEDLINE/PubMed, a 10% decline from the previous year. Document request traffic continues to decline in all Regions of the NN/LM due to expanded availability of electronic journals.

DOCLINE requests are routed to libraries automatically based on holdings data in its serial holdings database. At the end of FY2011, the holdings database contained over 1.6 million holdings statements for 60,334 serial titles held by 2,818 libraries.

NCBI, LO and staff at the Regional Medical Libraries continued to promote the use of PubMed's LinkOut for Libraries and "Outside Tool" as a means for libraries to customize PubMed to display their electronic and print holdings to their users. The number of libraries participating in LinkOut and Outside Tool increased to 833, 11% more than in the previous year; 2,636 libraries participate in LinkOut, a 9% increase over FY2010.

### *Reference and Customer Service*

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

NLM Customer Service responded to 93,296 inquiries from the public and libraries in FY2011 and answered 6,549 questions onsite in the Main Reading Room.

### **Outreach and Training**

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

professional meetings and publish articles to highlight NLM programs and services.

### *National Network of Libraries of Medicine*

The National Network of Libraries of Medicine (NN/LM) works to provide timely, convenient access to biomedical and health information resources for US health professionals, researchers, educators, and the general public. It is the core component of NLM's outreach program and its efforts to reduce health disparities and improve health information literacy. The network includes 5,942 full and affiliate members. Full members are libraries with health sciences collections, primarily in hospitals and academic medical centers. Affiliate members include some small hospitals, public libraries, and community organizations that provide health information services, but may have little or no collection of health sciences literature.

In addition to the basic NN/LM contracts, NLM funds contracts for three national centers that serve the entire network. The activities of one of these centers, the National Library of Medicine Training Center (NTC) located at the University of Utah, are described elsewhere in this report. The Outreach Evaluation Resource Center (OERC) at the University of Washington provides training and consulting services throughout the NN/LM and assists members in designing methods for measuring the effectiveness of overall network programs and individual outreach projects. In FY2011, the Center focused on an evaluation of the NN/LM National Emergency Preparedness and Response (NEP&R) initiative. The NEP&R evaluation showed strong and consistent evidence of increased emergency preparedness activity among NN/LM Network members from 2008 to 2011. The OERC also conducted an evaluation of the RML use of a Web conferencing tool, and worked with specific RMLs to conduct focused regional program evaluations. The Web Services Technology Operations Center (Web-STOC), also based at the University of Washington, provides ongoing technical management of the NN/LM Web sites and investigates, recommends, and directs the implementation of additional Web technology for teleconferencing, Web broadcasting, distance education, online surveys, etc. The Web-STOC began preparation for the NN/LM migration to an open source content management system, Drupal. The RMLs and other network members conduct many special projects to reach underserved health professionals and to improve the public's access to high quality health information. Most of these projects involve partnerships between health sciences libraries and other organizations, including public libraries, public health departments, professional associations, schools, churches, and other community based organizations. Some projects are identified by individual RMLs through regional solicitations or ongoing interactions with regional institutions; others are identified by periodic national solicitations for outreach proposals

issued simultaneously in all NN/LM regions. In all, the NN/LM funded 186 outreach projects in FY2011. Many of these projects focused on improving access to information for health professionals, consumers and the public health workforce.

With the assistance of other NN/LM members, the RMLs conduct 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 meetings of the Medical Library Association, the American Library Association, some of the health professional and library meetings in the Washington, DC area, and some distant meetings focused on health services research, public health, and history of medicine. In FY2011, NLM and NN/LM services were exhibited at 433 national, regional, and state meetings across the US.

#### *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 (CDC), 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, and SIS, serve on the Steering Committee, as do representatives from several RMLs.

The Partners Web site (PHPartners.org), managed by NLM with assistance from the New England RML, provides unified access to public health information resources produced by all members of the Partnership, as well as other reputable organizations. In FY2011, the Web site expanded significantly with more than 731 new links. Both visits, at 4%, and unique visitors, 8%, to the Web site continued to increase in FY2011 compared the prior year. In response to user feedback to improve the search functionality the search feature was modified to search sites of Partners organizations as well as the records underlying the PHPartners Web site itself.

One of the most popular resources on the PHPartners site has been pre-formulated PubMed search queries in support of the HHS Healthy People 2010 (HP 2020) initiative. In FY2010, NICHSR worked with the HHS Office of Disease Prevention and Health Promotion, to arrange for the development of comparable structured searches for Healthy People 2020, which launched at the beginning of FY2011. In FY2011, NICHSR worked with librarians and public health subject matter experts to develop “structured evidence queries” or “SEQs” for each of the Healthy People 2020 objectives and sub-objectives. By the end of the year, SEQs for 19 HP2020 topics had

been posted, with work underway for release of the remaining SEQs relatively early in FY2012.

#### *Special NLM Outreach Initiatives*

LO participates in many NLM-wide efforts to expand outreach and services to the general public and to address racial and ethnic disparities and participates in the Committee on Outreach, Consumer Health, and Health Disparities. For the seventh year, LO has worked in collaboration with NLM’s Director of International Programs to improve health information capacity in sub-Saharan Africa. When possible, a slot in the NLM Associate Fellowship Program is given to an African fellow. This year the program hosted Salima M’seffar, a librarian from the National Institute of Hygiene in Morocco.

For the past several years, LO coordinated a project to build journal capacity and enhance the quality of African medical journals by establishing partnerships between the editors of established medical journals and the editors of African medical journals. These four journals have been accepted for MEDLINE indexing, a major goal of the project: *African Health Sciences*, *Ghana Medical Journal*, *Malawi Medical Journal*, and *Mali Medical*.

#### *Historical Exhibitions and Programs*

HMD directs the development and installation of major historical exhibitions at NLM, with assistance from LHCBC, the Office of the Director, and OCCS. 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.

In FY2011, HMD helped to mark the 175<sup>th</sup> anniversary of the NLM through several initiatives, including preparing for release:

- *Hidden Treasure*, an illustrated book that showcases rare, beautiful, idiosyncratic, and surprising works in the collection. Each “treasure” is accompanied by a brief essay, written by a scholar, collector, artist, or physician. Each author wrote the texts of *Hidden Treasure* for the college-educated non-specialist reader — although the book will also be enjoyed by middle and high school students, health professionals and advanced scholars;
- *Building a National Medical Library on a Shoestring: 1872, the First Year*, a Web exhibition that focuses on the year when the Library, under the stewardship of John Shaw Billings, embarked on acquiring the most complete set possible of medical books and journals, an effort that set the course for the Library’s modern identity as the world’s largest medical library; and
- *In His Own Words: Martin Cummings and the NLM*, an online edition of selected speeches and articles of its former Director, Martin M. Cummings, MD.

Through FY2011, preparations continued for the opening of *Native Voices: Native Peoples' Concepts of Health and Illness*, which explores the interconnectedness of wellness, illness, and cultural life for Native Americans, Alaska Natives, and Native Hawaiians. This exhibition allows visitors to discover how Native concepts of health and illness are closely tied to the concepts of community, spirit, and the land. While preparing *Native Voices*, the Exhibition Program worked closely with NLM facilities staff to design and construct a new ceiling structure situated in the Rotunda gallery, which provides some protection for objects, as well as power, digital connections, and safety systems to the exhibition space.

The Exhibition Program continued to expand its education resources launching higher education modules for *Frankenstein: Penetrating the Secrets of Nature* and *Life and Limb: The Toll of the American Civil War*, and developing lesson plans related to Public Health posters from China. The Exhibition Program launched two traveling banner exhibitions: *The Henkel Physicians: A Family's Life in Letters*, about a Civil War-era family of physicians who lived in Virginia, based on an HMD manuscript collection, and *A Voyage to Health*, an exhibition that explores the recent revival of the ancient arts of navigation and voyaging of the people of Hawaii. In FY2011 the Exhibition Program supported tours of 25 copies of 11 different traveling exhibition titles. Over the last 10 years, exhibitions have traveled to 383 institutions in 46 states, the District of Columbia, and several international locations including Guam, Canada, Turkey, France, and Argentina.

The Exhibition Program partnered with other organizations to reach audiences beyond NLM's doors. *So Much Need of Service: The Diary of a Civil War Nurse* was an exhibition developed in conjunction with the National Museum of American History and featured in their public Document gallery in the summer of 2011. In August 2011, with the US Department of State, the Exhibition Program launched a pilot project to travel *A Voyage to Health* to new audiences in Argentina. In September, the Exhibition Program partnered with AMEDD Center of History and Heritage, part of the US Army Medical Command to travel *Life and Limb* to US Army audiences here and in Europe. In addition, the Exhibition Program initiated a working relationship with the Mount Vernon Estate and Gardens to develop a traveling exhibition about George Washington and health during the Colonial era.

The Exhibition Program also developed and produced special displays in the HMD reading room. The outstanding display this year was *From Craft to Profession: The Transition from Horse Farrier to Veterinarian*, which helped to mark World Veterinary Year while featuring NLM's international collection of rare horse veterinary books and manuscripts.

Additional HMD achievements included enhancing NLM Turning the Pages with its first book from NLM's collection of Islamic manuscripts, a 1537

manuscript copy of al-Qazwini's *Wonders of Creation*. This manuscript contains colorful illustrations and texts from the 13<sup>th</sup> century on plants, animals, and astronomical constellations. Turning the Pages also launched its first iPad application developed by LHNBCB's Communications Engineering.

This year NLM released two new Profiles in Science Web sites, for Charles R. Drew, MD, who has been called the "father of the blood bank," and for pioneering cardiac surgeon Clarence Dennis, MD. Charles Drew (1904-1950), while best known for the blood bank work, devoted much of his career to raising the standards of African American medical education at Howard University, and worked to break through the barriers that segregation imposed on black physicians. The Drew project is a collaboration with the Moorland-Spingarn Research Center at Howard University, the repository of his papers. Clarence Dennis (1909-2005) invented one of the first heart-lung bypass machines, and in 1951 was the first to use it to perform open-heart surgery. He was also an outstanding medical educator, first at the University of Minnesota, then at SUNY Downstate Medical Center, and finally at SUNY Stony Brook. Marking its thirteenth year in promoting research and teaching in the history of biomedical science, Profiles in Science program has to date digitized more than 175,000 pages of material, including published and unpublished materials ranging from books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audiotapes, and video clips.

HMD also began to integrate its seminar series more fully with its programs and exhibitions to reach and serve a wider audience. Complementing this initiative was staff presenting historical papers at professional and public meetings and publishing the results of their research in peer-reviewed publications and in blogs and new media outlets, such as the History News Network, which has 11,000 weekly subscribers.

#### *Training and Recruitment of Health Sciences Librarians*

LO develops online training programs in the use of MEDLINE/PubMed and other databases for health sciences librarians and other search intermediaries; oversees the activities of the National Network of Libraries of Medicine awarded this year to the University of Utah Health Sciences Library; directs the NLM Associate Fellowship program for post-Masters librarians; and develops and presents continuing education programs for librarians and others in health services research, public health, the UMLS resources, and other topics. LO collaborates with the Medical Library Association, the Association of Academic Health Sciences Libraries, and the Association of Research Libraries to increase the diversity of those entering the profession, to provide leadership development opportunities, to promote multi-institution evaluation of library services, and to encourage specialist roles for health sciences librarians. During this fiscal year, NLM worked with MLA to merge the National

Training Center Educational Clearinghouse with the MLA Educational Clearinghouse, reducing duplication and providing one comprehensive training resource.

In FY2011, the MEDLARS Management Section (MMS) and the NTC trained 879 students in 58 classes covering PubMed, the NLM Gateway, ClinicalTrials.gov, TOXNET, and the UMLS. An average of 13,407 unique visitors used the PubMed Tutorial for approximately 144,000 page views each month.

LO continues to create and enhance distance education resources. There are now 56 Quick Tours covering products such as PubMed, NLM Catalog, LinkOut for Libraries, and the UMLS. NLM and the NN/LM are increasing the use of Webcasts to provide current information on NLM's products and services as a cost effective and timely way to reach larger numbers of users. NICHSR continues to make available its suite of courses on health services research, public health, and health policy. In FY2011, in addition to the two-part Health Technology Assessment Webinar series, mentioned earlier, NICHSR sponsored a joint Academy Health-MLA Webinar series on Grey Literature.

The NLM Associate Fellowship program had nine participants in FY2011, with six first-year Fellows on site at NLM and three second-year Fellows at other

libraries. LO placed three second year fellows at Yale University, University of Maryland, and the NIH Library. In September 2011, a new group of four first year Fellows began their year at NLM and five associate fellows began their second-year at health sciences libraries throughout the US. NLM works with several organizations on librarian recruitment and leadership development initiatives. Individuals from minority groups continue to be underrepresented in the library profession. LO provided support for scholarships for minority students through the American Library Association, Medical Library Association, and the Association for Research Libraries (ARL). In the summer of 2011, NLM participated for the third year in the ARL Career Enhancement Program (CEP), supported by an Institute of Museum and Library Services (IMLS) grant, and designed to attract minority library school students to careers in research libraries. NLM hosted three ARL CEP students during the summer. LO also supports the NLM/AAHSL Leadership Development Program, which provides leadership training, mentorship, and site visits to the mentor's institution for an annual cohort of five mid-career health sciences librarians. These programs support the development of a strong and diverse professional biomedical library workforce for the future.

**Table 1: Collections**

<i>Physical</i>	<i>Total<sup>1</sup></i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
Monographs <sup>2</sup>				
Before 1500	597	1	1	0
1501-1600	6,053	7	21	7
1601-1700	10,343	5	24	4
1701-1800	272,727	21	54	17
1801-1870	256,653	66	108	31
1871-Present	847,895	14,650	12,855	14,917
Bound Serial Volumes <sup>3</sup>	1,434,815	21,924	23,418	21,910
Microforms <sup>4</sup>	606,066	15	249	130
Audiovisuals and Computer Software	89,950	1,224	1,335	994
Prints and Photographs	69,316	20	11	0
Manuscripts <sup>5</sup>	15,658,457	1,143,275	3,530,800	2,448,075
Withdrawn Items	(131,568)	(1,272)	(1,253)	(3,304)
<b>Total items</b>	<b>19,121,304</b>	<b>1,179,936</b>	<b>3,567,623</b>	<b>2,482,781</b>
<i>Digital</i>	<i>Total</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
PubMed Central Articles	2,277,749	223,404	179,153	191,087
PubMed Central Titles <sup>6</sup>	897	203	158	164
Bookshelf Titles <sup>7</sup>	1073	295	501	59
Digital Collections Repository <sup>8</sup>				
Texts <sup>9</sup>	1904	1406	518	NA
Films	29	18	11	NA

<sup>1</sup> Total: Numbers are cumulative as of the end of the fiscal year.

<sup>2</sup> Monographs: A bibliographic resource complete in one part or finite number of separate parts. Includes Americana, theses and pamphlets. Starting in FY2011 numbers for these materials are reported under monographs by publication year.

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

<sup>4</sup> Microforms: Reduced size reproductions of monographs and serials including microfilm and microfiche.

<sup>5</sup> Manuscripts: Total manuscripts equivalent to 8,948 linear feet

<sup>6</sup> PMC titles: Only fully deposited titles

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

<sup>8</sup> Digital Collection Repository: Digitized content in the public domain. In the future will contain born digital items as well as reformatted items.

<sup>9</sup> Includes monographs and serials such as annual reports.

**Table 2: Collection Activities**

<i>Acquisitions and Processing</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
Active Serial Subscriptions	19,731	20,645*	20,096
Publications Processed <sup>10</sup>			
Modern Manuscripts <sup>11</sup>	866 (in ft)	681 (in ft)*	2776 (in ft)
Serial Pieces	117,091	119,067	129,053
Monographs (1914- ), Audiovisuals, etc.	20,639	21,194	23,224
Monographs (pre-1914) <sup>12</sup>	886	1,041*	592
Prints and Photographs	897	1,845	3,914
<b>Total</b>	159,244	163,486	176,879
<i>Obligations</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
Publications	\$10,516,299	\$11,222,292	\$9,549,598
(For Rare Books)	(\$299,794)	(\$299,822)	(\$294,219)
<i>Preservation</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
Volumes Bound	18,997	19,629	18,866
Volumes Repaired Onsite <sup>13</sup>	2,304	2,727	3,174
Audiovisuals Preserved	594	414*	268
Historical Volumes Conserved	340	4,125	185
Pages Digitized	787,865	72,622	328

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

<sup>10</sup> Publications processed: Serial issues, monographs and nonprint receipts processed

<sup>11</sup> Modern manuscripts: not included in total

<sup>12</sup> Monographs (pre-1914) includes historical manuscripts

<sup>13</sup> Volumes repaired onsite: General Collection monographs and serials only

**Table 3: Cataloging and Indexing**

---

<i>Cataloging</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
General Collection Items <sup>14</sup>	21,268	18,820	20,615
Historical Monographs (pre-1914)	1,128	4,328	4,675
Modern Manuscripts	843 (in ft)	2,018 (in ft)	1,437 (in ft)
Prints and Photographs	508	14,784*	31,605
Historical Audiovisuals	105	792	2,700
<i>Indexing</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
Citations Indexed for MEDLINE	724,831	699,420	712,675
Journals Indexed for MEDLINE	5,559	5,484	5,394

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

---

<sup>14</sup> Items: Includes monographs, serials, nonprint and integrating resources.

**Table 4: Services to the Public**

<i>Document Delivery</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
Interlibrary Loan Requests Received	244,701	256,459	271,455
Interlibrary Loan Requests Filled	200,581	213,330	229,577
General Reading Room Requests Received	119,209	145,240	170,919
General Reading Room Requests Filled	102,054	124,815	146,189
History of Medicine Reading Room Requests Filled	23,724	14,843	13,439
<i>Customer Service Inquiries</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
Offsite Inquiries <sup>15</sup>			
General	97,796	91,540	83,140
History of Medicine	8,667	8,811	8,567
Onsite Inquiries <sup>16</sup>			
General	6,531	7,076	7,200
History of Medicine	15,057	14,635	13,248
<i>Data Licensees</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
MEDLINE	688	621	532
UMLS	6,566	6,116	4,853
<i>Tours and Visitors</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
Exhibitions			
Visitors	2,624	7,183	N/A
Daily Tours			
Tours	100	115	112
Visitors	548	622	646
Special Tours			
Tours	71	69	66
Visitors	1,392	1,558	2,154

<sup>15</sup> Offsite Inquiries: inquiries via telephone, fax, US mail, and e-mail. Includes BSD interactions with data licensees.

<sup>16</sup> Onsite Inquiries: in person

**Table 5: Selected Web Resources**

<i>Resource</i>	<i>FY2011</i>	<i>FY2010</i>	<i>FY2009</i>
<b>ClinicalTrials.gov</b>			
Number of Trials	118,682	100,919	83,540
Page Views <sup>17</sup>	1,347,564,098	857,270,086	785,589,163
Visitors <sup>18</sup>	9,518,503	10,583,773	10,794,930
<b>DailyMed</b>			
Number of Labels	31,000	11,600	5,700
Page Views	118,163,000	70,195,000	66,546,000
Visitors	5,166,000	3,035,000	1,902,000
<b>Genetics Home Reference</b>			
Summaries	1622	1408	1153
Page Views	142,967,488	118,267,551	36,465,215
Visitors	4,819,287	3,641,936	3,005,955
<b>Household Products Database</b>			
Number of Products	11,000	10,000	9,000
Page Views	10,254,423	10,984,725	140,340,909 <sup>19</sup>
Visitors	694,042	850,332	875,237
<b>MEDLINE/PubMed</b>			
PubMed Citations	21,200,952	20,209,357	19,181,741
Page Views	4,783,804,040	4,104,309,671	3,792,907,905
Unique Visits <sup>20</sup>	539,184,717	366,396,717	349,191,616
Searches <sup>21</sup>	1,834,537,403	1,578,714,477	1,281,180,957
<b>MedlinePlus</b>			
Number of Topics (English/Spanish)	903/881	859/826	805/762
Page Views	728,000,000	716,000,000	660,000,000
Visitors	164,000,000	153,000,000	123,000,000
<b>NLM Home Page</b>			
Page Views	39,300,000	44,500,000	42,700,000
Visitors	8,300,000	7,700,000	7,900,000
<b>ToxTown</b>			
Page Views	6,164,031	6,096,928	6,352,651
Visitors	192,737	286,004	300,230

<sup>17</sup> Page Views: Number of times that a single page is viewed or downloaded.

<sup>18</sup> Visitors: Number of people visiting a Web site in a defined period of time.

<sup>19</sup> In 2009, Household Products was hit by a "hacking site" causing repetitive hits equaling tens of thousands per day.

<sup>20</sup> Unique Visits: Total number of times that all users visit a Web site, regardless of the number of individual pages viewed.

<sup>21</sup> Searches: Number of searches performed.

# SPECIALIZED INFORMATION SERVICES

*Steven Phillips, MD*  
*Associate Director*

The Division of Specialized Information Services (SIS) at the National Library of Medicine (NLM) offers a diverse collection of publicly accessible, free online information resources and services in toxicology, environmental health, chemistry, HIV/AIDS, disaster management, minority health, and other specialized topics. SIS includes an Outreach and Special Populations Branch, which actively seeks to improve access to high quality health information by underserved and other targeted populations. SIS also manages NLM's Disaster Information Management Research Center (DIMRC).

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 used the most up-to-date and relevant information and technologies to provide rapid and effective access to the latest toxicological and environmental health information. We not only create databases ourselves, but direct users to relevant sources of toxicological and environmental health information wherever these sources may reside as part of the Library's role in collecting and organizing health and medical information. Identifying high-quality resources is 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 new guides on Developing and Using Medicines for Children, Epigenomics, Water Pollution, and Hexavalent Chromium and other Chromium Compounds.

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

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

This year, DIMRC continued to expand its activities as well as respond to international crises, such as the earthquake, tsunami and radiation incident in Japan. Sections of the Radiation Emergency Medical Management (REMM) tool were translated into Japanese

to provide guidance to health professionals in Japan. The Chemical Hazards Emergency Medical Management (CHEMM) resource, a companion to REMM, was released in FY2011. This tool assists emergency managers, planners, responders and other healthcare professionals in responding to mass casualty chemical events.

In FY2011, SIS completed its third year of co-directing the Public Health Informatics course at the Uniformed Services University of the Health Sciences (USUHS). This course is jointly sponsored by the USU Biomedical Informatics Department and NLM to provide students with a conceptual framework for understanding the emerging field of public health informatics, highlighting the latest approaches and technologies used by NLM and other organizations. New for 2011 were speakers from the National Aeronautics and Space Administration (NASA) and the Centers for Disease Control and Prevention (CDC).

The SIS Web site provides a view of the full range of the varied programs, activities, and services of the Division. Although users typically enter through a specific entry point for the topic of interest (TEHIP, HIV/AIDS, disaster information, or minority health), the Divisional Web site (<http://sis.nlm.nih.gov>) includes program descriptions and documentation. Continuous refinements and additions to our Web-based systems are made to allow easy access to the wide range of information collected by this Division. In addition to the Web sites and databases, SIS added Twitter and Facebook pages to its repertoire of communication tools, developed apps for smartphones, and used crowd-sourcing for resource development.

## Toxicology and Environmental Health Resources

NLM SIS (@NLM\_SIS) 9/13/11 5:12 PM RT  
[@publichealth](#): FDA: 'Apple juice is safe to drink.' FDA explains relationship btwn arsenic & apple juice:  
[goo.gl/gPxVU](http://goo.gl/gPxVU)

Here is an example of our new way to reach people with information on toxicology and environmental health. This tweet went out from our Specialized Information Services (SIS) Twitter account. We have moved beyond the Web resources we provide in the area of toxicology and environmental health information to social media methods that allow us to rapidly deploy new important relevant information on our own resources or other ones as well. We also use a toxicology and environmental health listserv, Facebook accounts, and RSS feeds to supplement our traditional Web site and searchable databases. This helps us reach the general public as well as scientists, researchers, and health workers. In addition, we are realizing that the trend in information access is mobile, and we are in process of developing or enhancing applications for mobile access.

Our TOXNET (TOXicology Data NETwork) <http://toxnet.nlm.nih.gov> databases continue to be the core content in our information resources. These databases are

varied in what they cover, and how they are created. Some are legacy ones, evolving from their earlier forms as we included information on a wider range of toxic chemicals, such as adding Superfund cited compounds, or taking on the possible safety issues with nanomaterials. Enhancements to TOXNET continue, based on user feedback and upgrades or additions of data and capabilities. The following TOXNET databases underwent significant expansion and development in FY2011:

- **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. This year we added the first group of complementary and alternatives medicines (CAM) to LactMed. These records focus on supplements that are frequently used by nursing mothers.
- **HSDB** (Hazardous Substances Data Bank), a peer reviewed database focusing on the toxicology of over 5,500 potentially hazardous chemicals. This flagship database continues to be enhanced with records on nanomaterials. This year, structures from ChemIDplus were added to HSDB records and are displayed as images at the start of the record. We prepared for the re-engineering of the HSDB to better incorporate new resources important to today's toxicologists by planning an HSDB Needs Assessment project.
- **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 nearly 4.6 million citations, almost all with abstracts and/or index terms and CAS Registry Numbers. This year, we added updated records on toxicology research projects from the CRISP database.
- **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 384,000 chemical records, of which over 300,000 include chemical structures. ChemIDplus includes some toxicity data as well as locators to many important national and international listings of chemicals. This year ChemIDplus structures were added to the display of several TOXNET databases.
- **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, and

the first steps were taken to add information about nanomaterials in household products. Also, a memorandum of understanding was signed with the Consumer Product Safety Commission (CPSC), to discuss possible types of nanomaterials and additional categories of products to add to the Household Product Database.

- **Haz-Map**, 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 Web portal providing information on resources about alternatives to the use of live vertebrates in biomedical research and testing, developed as part of NLM's participation in the Interagency Coordinating Committee on the Validation of Alternate Methods (ICCVAM). Work was nearly finished this year to update and expand the content of the ALTBIB resource to align with the fast moving development, evaluation, and validation of alternative methods.  
In addition to the core TOXNET databases, SIS supports many other environmental health databases and resources:
- **TOXMAP**, a Geographic Information System (GIS) that uses maps of the United States to help users visually view data about chemicals released into the environment and easily connect to related environmental health information. Enhancements released in FY2011 include EPA Toxics Releases (2009 data) and EPA Superfund National Priorities List (NPL) data. Additional classroom exercises were released in coordination with the California University of Pennsylvania. A map of current US nuclear power plants was created with a TOXMAP template. Work was continued on a new version of TOXMAP which will more closely align with the look and feel of current mapping tool.
- **Drug Information Portal**, provides current drug information for over 28,000 drugs with links to many credible additional online resources. During FY2011, changes were made to the search interface to allow easier selection of drug names 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 appearance information. This year, we continued to work with FDA on the development of

this resource, which is an integral part of the patient-safety initiative with the FDA. 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. The images produced through this project are now included in the Pillbox Web site and Web service.

- **Dietary Supplements Labels Database**, was enhanced with additional labels from the many brands available in the marketplace. SIS is also collaborating with the NIH Office of Dietary Supplements (ODS) in developing a comprehensive labels database intended for researchers. The ODS prototype label database will be developed during the next year.
- **Comparative Toxicogenomics Database (CTD)**, developed at the Mount Desert Island Biological Laboratory and included in TOXNET searching. CTD receives funding from the National Institute of Environmental Health Sciences (NIEHS) and NLM. This database focuses on molecular mechanisms by which environmental chemicals affect human disease, providing insight into complex chemical-gene and protein interaction networks.
- **ToxLearn**, a tutorial providing an introduction to concepts and principles of toxicology, is under development as a joint project with the Society of Toxicology. Module 1, Introduction to Toxicology and Dose-Response, was released this year. Future modules will be released as they are developed.

### Outreach Initiatives

#### *The Environmental Health Information Partnership (EnHIP)*

Enhances the capacity of minority serving academic institutions to reduce health disparities through the access, use and delivery of environmental health information on the campuses of HBCUs, tribal colleges and minority serving institutions and in their communities. EnHIP celebrated its 20<sup>th</sup> anniversary with a successful meeting of its 22 partnering institutions including 15 HBCUs, three tribal colleges, three Hispanic-serving institutions and one Alaskan university. The program, *From TIOP to EnHIP: Evolution of Environmental Justice*, featured a 20-year retrospective by Melvin L. Spann, PhD, who started the TIOP program in 1991 with nine member institutions. Robert Bullard, PhD, from the Environmental Justice Resource Center, and Robert C. Warren, DDS, MPH, DrPH, MDiv, director of the National Center for Bioethics in Research at Tuskegee University, provided a history of the environmental justice movement. Presentations also featured Dr. Steven Phillips, MD on NLM's role in Haiti's disaster recovery as well as principal investigators of

EnHIP funded projects. Thirteen small outreach projects were funded with EnHIP partner institutions in FY2011.

#### *United Negro College Fund Special Programs (UNCFSP)/NLM HBCU Access Project*

Continued its partnership to work with HBCUs to incorporate NLM information resources into HBCU courses and other campus and community activities. In June, UNCFSP and NLM held its annual meeting *Celebrating 10 Years of Health Information ACCESS at HBCUs*. This milestone meeting highlighted the accomplishments of the 10-year partnership and project funding. Featured speakers included Frances Goins-Ashe, RN, MPH, deputy director, Office of Women's Health, DHHS, and William Coleman, Jr. PhD, scientific director, NIMHD. Other topical areas included health informatics, health literacy, social media and outreach project reports. Following the meeting, two additional sessions were offered, an online training workshop held at NLM and a technical assistance workshop at UNCFSP on developing outreach proposals. During FY2011, four HBCUs were funded for community or campus outreach projects.

#### *Outreach to Minority Health Professionals*

SIS continued its efforts to increase the awareness of and use of NLM online resources by minority health professionals. Hands-on training, presentations and demonstrations were conducted at the Student National Medical Association Annual Medical Education Conference, National Medical Association, National Black Nurses Association, and at several regional conferences.

In FY2011, SIS staff expanded its work with Black Greek Letter Organizations. The organizations are involved actively in local outreach efforts to combat the health disparities that adversely affect minorities. Staff worked with Delta Sigma Theta Sorority, Alpha Kappa Alpha Sorority and Alpha Phi Alpha Fraternity to introduce NLM resources to promote health literacy and health awareness.

#### *Outreach to Librarians*

SIS worked with NLM Associate Fellow Stephen Kiyoi on his project, *Outreach to Social Workers*. The project involved an investigation into the social work profession and their need for and use of biomedical and consumer health information. Research included a literature review, interviews with librarians, and a pilot survey with NN/LM office staff. Results of the research were presented to NLM senior staff in August. Paper and poster submissions were made for the 2012 Medical Library Association annual meeting.

NLM continues to support the training and education of minority librarians through funding of two graduate assistantships for students in the Knowledge River Program at the School of Library and Information

Science, University of Arizona. These students work in the Health Sciences Library and are encouraged to pursue careers as health sciences librarians.

### Special Population Web Sites

SIS began a project to develop digital stories with American Indian community members for **American Indian Health Web Portal**. Four training sessions were held in FY2011 with participants from the Fond du Lac Human Services, Inter-Tribal Council of Michigan, Shoshone Women's Society, and Montana Wyoming Tribal Leaders Council. Each participant received training in digital storytelling methods and instruction on the use equipment and software to prepare their personal story. Some of the stories will be available on the American Indian Health Web Portal. A poster presentation on the digital storytelling project was presented at the Medical Library Association meeting. Two new sections were added to the portal in FY2011: "Exercising and Keeping Fit" (<http://americanindianhealth.nlm.nih.gov/exercise.html>) and "Eating Healthy and Nutrition" (<http://americanindianhealth.nlm.nih.gov/eating.html>).

Working with our partner, the Association of Asian Pacific Community Health Organizations (AAPCHO), SIS began adding more content for the Complementary and Alternative Medicine section of **Asian American Health** (<http://asianamericanhealth.nlm.nih.gov/>). AAPCHO is conducting focused outreach with its members that serve Southeast Asian and South Asian populations as well as with additional community health centers staff to promote the site as well as gather more CAM resources.

WomensHealthNIH Sep 13, 2011, 9:10 am via HootSuite #Smoking Might Raise #Sex Hormone Levels After #Menopause <a href="http://ow.ly/6s7Dy">http://ow.ly/6s7Dy</a>
--

Smoking Might Raise Sex Hormone Levels After Menopause

**Women's Health Resources (WHR)**, a collaboration with the Office of Research of Women's Health, NIH, continued to expand research dissemination through social media. Facebook (see <http://www.facebook.com/WomensHealthResources>) has 1,158 "likes" and Twitter (see <http://twitter.com/WomensHealthNIH>) grew to over 11,120 followers. A presentation on WHR social media initiatives was provided during the *Women's Health and Social Media* forum held at NIH in March. With funding from the Office of Research on Women's Health, an outreach program for researchers and consumers is under development. Focus will be on increasing the capacity for sex differences research in junior and senior researchers at universities and increasing female consumers' knowledge and use of research information.

SIS is a partner in the **Refugee Health Information Network (RHIN)**, a national collaborative partnership of the Association of Refugee Health Coordinators, NLM and the Center for Public Service Communication (CPSC.) RHIN is committed to providing quality multilingual, 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. This year, the RHIN Web site was redesigned and debuted at the end of February 2011.

### Social Media

Social media offers many avenues to reach new audiences. SIS expanded its use of social media for promotion and outreach this year. SIS launched the **NLM4CareGivers** pilot project to learn more about the use of social media as a viable option to deliver health information. During the pilot, SIS will increase the awareness of NLM's resources among groups that are active users of online health information. Family caregivers were selected for this pilot project due to their high use of the Internet and their use of blogs and other social media outlets for health information. Facebook ([http://facebook.com/NLM\\_4Caregivers](http://facebook.com/NLM_4Caregivers)) and Twitter (@NLM\_4Caregivers) accounts were established and metrics will be collected from each to determine success.

There are now Facebook accounts for several SIS resources such as NLM 4Caregivers, Women's Health Resources and *AIDSinfo*. In addition, Twitter feeds are available for several of SIS resources including:

@NLM\_SIS  
@NLM\_OSP  
@NLM\_DIMRC  
@NLM4Caregivers  
@AIDSInfo  
@refugeehealth

### SIS K-12 Initiatives

#### *Tox Town*

Was enhanced with new content in English and Spanish in the neighborhoods of Tox Town, Tox City, Tox Farm, Tox Port, and US Mexico Border scene. A funeral home site was added to Tox City and a new Southwestern scene is under development with assistance from Diné College. The Southwest scene will introduce environmental hazards such as uranium tailings, abandoned mines, coal-fired power plants, dust storms and more.

The K-12 team continued work on the curriculum for the Tox Town-based environmental health afterschool club curriculum for middle school students. The team reviewed and edited lesson plans developed in FY2010 in collaboration with a group of middle school teachers. The

expansion involved developing additional activities and working with editors to streamline the style and the format. The resulting curriculum, consisting of six three-to-four lesson units, will be released as a set of PDF books in FY2012. An adapted week-long version of one of the curriculum units was pilot tested in a high school environmental science classroom in Washington, DC.

*Tox Mystery*, an interactive Web site for children ages 7-10, was released at the end of FY2006. The game-like interface introduces Toxie the Cat who helps children explore potential chemical hazards in the house. Outreach activities this year included Toxie the Cat Day at the Portland Children's Museum, Maine with children participating in Tox Mystery related activities. SIS also conducted focus groups for a usability evaluation of Tox Mystery and developed a plan for expanding the site's content depth with additional information about the chemicals and hazards described in the game.

### **Environmental Health Student Portal**

Is a Web site for middle school students and teachers, developed in FY2010. In FY2011, we obtained set-aside evaluation funding from the NIH Evaluation Office and initiated a usability study of the Portal.

SIS began development of **GeneEd**, a Web site that will provide a single repository for genetics research, study guides, lesson plans, experiments and activities for teachers and students in grades 10-12. The concept for the site came from interest expressed by educators at a genetics conference and confirmed by a survey distributed by the National Association of Biology Teachers. The Web site will provide a single, safe source for teachers to use to introduce, teach, or supplement existing genetics course materials and to recommend to students. SIS will lead the development of the site and will seek input from groups that specialize in genetics content and research, including LHCBC Genetics Home Reference staff, NCBI, and the NIH Genetics and Rare Diseases Information Center.

In FY2011, members of the K-12 team participated as exhibitors and presenters in a number of conferences, including National Science Teacher Association conference and International Conference of Science in Society and wrote and submitted several manuscripts describing the team's educational and research activities. An example of such publication is an article for *Science Teacher*, the membership journal of the National Science Teacher Association. The article, describing using Tox Town to support science classroom argumentation, will be published in the summer of 2012.

SIS continued its participation in the NLM distance learning program. Three schools participate in the program: King Drew Medical Magnet High School in Los Angeles, CA, Kotzebue High School, Kotzebue, AK, and Farrington High School in Honolulu, HI. The theme of this year's programming was global health. Sessions were held twice monthly during the school year with various

speakers. SIS and NLM staff made a site visit to Kotzebue High School in September to discuss technical issues as well as to speak with past students about the impact of participating in the distance learning program. In addition, NLM hosted seven teachers from the high schools to discuss future programming and to update them about cloud computing, History of Medicine programs, the Office of Science Education, and more.

As part of the NLM Adopt-a-School Program, a partnership with the NLM Office of the Director, SIS coordinated summer internships for five students from local area high schools. SIS hosted a student from Woodrow Wilson High School who assisted with the development of 20 mini-guides to SIS products. The other students were hosted by the Office of the Director, Lister Hill Center, OCCS, and Library Operations.

In March, SIS provided support for the 25th Association of Minority Health Professional Schools Annual Symposium in Newport News, Virginia. The meeting focused on local and regional pipeline programs that included enrichment activities to encourage and promote minority student interest in science and math. NLM introduced students to its online resources at an exhibit.

### **HIV/AIDS**

NLM has continued its HIV/AIDS-related outreach efforts to community-based organizations, patient advocacy groups, faith-based organizations, departments of health, and libraries. The program, begun in 1994, provides support to design local programs for improving information access for HIV/AIDS patients, the affected community, and their caregivers. Emphasis is on providing information or access in a way meaningful to the target community. Projects must involve one or more of the following information access categories: information retrieval, skills development, Internet access, resource development, social media or document access. In FY2011, NLM made awards to the following organizations: AIDS Educational Global Information System (AEGIS); Alliance of Border Collaboratives (ABC); Friendship House Association of American Indians, Inc.; George Washington University; Justice Resource Institute; Lee County (IL) Health Department; National Native AIDS Prevention Center; University of New Mexico; and Washington Area Consortium on HIV in Youth/Metro Teen AIDS.

NLM continues to manage the AIDSinfo program (*AIDSinfo/infoSIDA*), which serves as the dissemination point for the federal HIV/AIDS treatment guidelines, HIV/AIDS-related clinical trials information, as well as other treatment, research and prevention information. This program is supported by NLM, the NIH Office of AIDS Research, National Institute for Allergy and Infectious Diseases, the Centers for Disease Control and Prevention and the Health Resources and Services Administration. In FY2011, *AIDSinfo* had numerous technical upgrades

including the development (March 2011) of a database driven backend to support the delivery of the guidelines in a searchable HTML format as well as continuing the printer-friendly PDF. *AIDSinfo* also implemented the use of an offsite disaster recovery, failover hosting site in Portland, Oregon, to minimize interruptions in service due to power failures or local server problems. Within 30 seconds of a power outage or server failure, traffic to the *AIDSinfo* and *infoSIDA* Web sites will be automatically redirected to the Oregon-based disaster recovery server. In February 2011, *AIDSinfo* began using Google Urchin to capture Web metrics. An upgrade from the previously used WebTrends analytics software, the Google-based software offers *AIDSinfo* a cost-effective solution to more accurately track Web usage. *AIDSinfo* also supported multiple updates to the treatment guidelines during FY2011. The subject matter experts are working groups of the NIH Office of AIDS Research Advisory Council.

*AIDSinfo* developed several products to extend its reach to an expanding audience of mobile users: an application for the Apple iPhone for the *AIDSinfo* Glossary of HIV/AIDS-Related Terms (October 2010); a video highlighting the 10 key *AIDSinfo* informational services/resources, presented on NIH Office of the Director YouTube channel in observance of World AIDS Day (December 1, 2010); the *AIDSinfo* Facebook page (launched March 2011); and the *infoSIDA* mobile site (released June 2011). *AIDSinfo* revamped the Health Topics sections of both sites in March 2011 to provide easy access to links to government HIV/AIDS-related resources. Enhancements included an updated design, additional content, an A-Z index, a search feature, and improvements to help users easily navigate to resources in the sections.

### Mobile Technologies

SIS develops mobile apps to expand the dissemination of health information to emergent mobile platforms, such as iPhone/iPad, Android, and BlackBerry. In addition to those described in other sections of this chapter, new apps released in FY2011 include:

*Embryo*, a new app for the iPhone/iPad, was released this year following a collaboration between NLM and the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the Virtual Human Embryo Project (Louisiana State University) and the National Museum of Health and Medicine's Human Development Anatomy Center. The *Embryo* app provides a collection of digital serial sections of early stage human embryos. The *Embryo* app was released in March 2011.

*LactMed*, available as iPhone/iPad, and Android apps, provides information about the effects of drugs and dietary supplements when used during breastfeeding. It is a great mobile resource for nursing mothers, nurses, pharmacists,

lactation consultants, and pediatricians. *LactMed* apps were first released in June 2011.

*REMM*, previously available on iPhone/iPad, BlackBerry, and Windows Mobile, provides health care providers with guidance on the diagnosis and treatment of radiation injuries during radiation and nuclear emergencies. In April 2011, an Android version of Mobile *REMM* was released.

### Disaster Information

In May 2010, the NLM Board of Regents approved the establishment of an ad hoc Working Group on Disaster Health Information Management. The group was formed to review the current activities of NLM's Disaster Information Management Research Center (DIMRC) and provide guidance on future directions. The Working Group will assist in revising priorities and resource allocations/estimates to address how NLM can enhance its contributions to the nation's disaster information management agenda. The first meeting was held in May 2011.

The SIS Office of the Disaster Information Research Center (DIMRC) coordinates disaster health information management activities across NLM. Major initiatives in FY2011 included the launch of the Chemical Hazards Emergency Medical Management (CHEMM) tool, and the development and promotion of resources in response to the earthquake, tsunami and radiation incident in Japan. CHEMM is designed to assist emergency managers, planners, responders and other healthcare professionals in preparing for and responding to mass casualty chemical events. It provides just-in-time medical management guidance and includes the CHEMM-IST tool to assist in the identification of the type of chemical involved in an incident. CHEMM was developed by NLM in conjunction with the HHS Office of the Assistant Secretary for Preparedness and Response (ASPR), as well as with subject expertise provided by several other agencies, including the Department of Homeland Security.

SIS continued work on the Chemical Emergency Medical Management (CHEMM) tool, with CHEMM released to the public in June 2011. CHEMM provides just-in-time medical management guidance to health professionals as well as assistance with planning and training for mass casualty chemical events. CHEMM also includes tools for modeling exposure to chemicals and provides syndromic identification and treatment guidance. *REMM* and CHEMM are both produced in conjunction with HHS ASPR.

The Wireless Information System for Emergency Responders (WISER) is now available for the Android, in addition to the iPhone/iPod touch, BlackBerry, Windows mobile, Windows, and on the Web. WISER is a system designed to assist emergency responders in hazardous materials incidents and contains information on chemical, biological, and radiological agents. There are over

150,000 downloads of WISER to the various platforms. A new interactive Chemical Reactivity capability was incorporated into WISER for Windows, based on collaboration with Environmental Protection Agency (EPA) and The National Ocean and Atmospheric Association (NOAA). Additional substances were also introduced including oil spill response substances in the news. Radiation Emergency Medical Management (REMM), which provides guidance to health care providers about diagnosis and treatment of radiological/nuclear events, is also available for all of these same platforms.

A new design for the DIMRC Web site was launched early in FY2011. The new site has 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. The new look mirrors the look of CHEMM and the TEHIP Web sites. The former Resource Guide for Public Health Preparedness was renamed as the Resource Guide for Disaster Medicine and Public Health, which reflects its broader scope and coverage. Major enhancements improved and streamlined the management of this database. The redesigned database was launched along with the new DIMRC Web site.

#### *Disaster Information Outreach*

Significant progress was made in the Disaster Information Specialist program in FY2011. The highlight was the Disaster Information Outreach Symposium held March 29-30, 2011. This 1.5-day symposium brought together nearly 200 librarians, information professionals, and disaster medicine and public health professionals (in addition to over 150 via videocast). The keynote speaker, RADM Nicole Lurie, MD, Assistant Secretary for Preparedness and Response, HHS, praised the efforts of librarians and encouraged them to work in their communities as access to timely, authoritative information is essential in disasters. NLM also worked with the Medical Library Association to initiate the Disaster Information Certificate program, and the first course was taught at an MLA chapter meeting and online.

NLM issued the first request for proposals for disaster information outreach projects and funded seven projects in FY2011 designed to enhance collaboration and increase awareness and knowledge of disaster health information. The projects required an organization working in disasters (public health department, emergency management department, fire/rescue service, academia, etc.) to partner with a library to improve access and

knowledge of NLM and other disaster health resources and to identify roles for librarians in disaster management.

The Disaster Outreach Listserv continued to grow, from 450 subscribers at the end of FY 2010 to over 650 subscribers at the end of FY2011.

DIMRC continued to work with the Pan American Health Organization (PAHO) and the Regional Disaster Information for Latin America and the Caribbean (CRID) on the network of disaster information centers in the region. The first meeting of both the Central American and South American networks of Disaster Information Centers was held in December 2010 and the two networks voted to merge into one network serving Latin America (RELACIGER – Spanish acronym for the Latin American Network of Disaster Health Information Centers). In addition, an NLM-supported Toolkit for Disaster Information Centers was released. This toolkit assists organizations interested in creating a new information center in their community.

NLM and PAHO also supported CRID in the development of a Haiti Disaster Information Center Web site to ensure that vital health information is not lost following the earthquake and cholera outbreaks of 2010. To date, the Web site contains nearly 1,000 documents (situation reports, after action reports, assessments, etc.) related to health. CRID, PAHO, and NLM are working to identify a library or other organization in Haiti that can assume responsibility for the maintenance and development of the Web site.

#### *Research and Development*

The DIMRC office continued to coordinate and conduct a number of informatics and communications research and development projects on behalf of the Bethesda Hospitals' Emergency Preparedness Partnership (BHEPP). NLM joined BHEPP in 2008 to help with common issues at hospitals during emergencies: communication, patient management, family reunification, and information access. Three NLM divisions, SIS, OCCS, and LHNCBC are working on these projects. DIMRC staff have developed several projects, including a back-up digital communication system via the Military-Affiliated Radio System (MARS), and the development of a virtual world prototype for training hospital personnel in the nationally mandated Hospital Incident Command System. The digital MARS radio system was tested at the Collaborative Multi-Agency Exercise (CMAX) in October 2010. Information gathered at the exercise provided guidance for further development of the projects in FY2011. Further testing will occur at subsequent drills at the hospital.

### *DIMRC 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 bodies, including the NIH Biodefense Research Coordinating Committee, as well as 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.

### **Evaluation and Research**

In addition to evaluation activities described elsewhere in this report, SIS continued to use The American Customer Satisfaction Index (ACSI) to obtain user input on selected SIS resources. In March 2011, updated ACSI surveys were launched on both *AIDSinfo* and *infoSIDA* sites. Data were used to develop some of the recommendations for redesign of both these sites. Between November 2010 and January 2011, *AIDSinfo* conducted usability studies of the *AIDSinfo* and *infoSIDA Web sites*, which also helped inform the redesign process.

SIS also continued two evaluation activities that were initiated with FY2010 set-aside Evaluation Office funding. The first is evaluation of TOXNET, a collection of toxicology, environmental health, and chemical databases. The evaluation is now completed. The results are being used to inform ongoing redesign. The second is the development of an evaluation plan for NLM AIDS Community Information Outreach Program. This activity is being conducted by the Columbia University Department of Biomedical Informatics. The project started in FY2011, and will be completed in FY2012. We

anticipate continuing working with Columbia University to implement the evaluation plan.

SIS participated in a number of research activities in the fields of program evaluation, consumer health informatics, and science education. Continuing an evaluation research project that was initiated in FY2010, we compiled and analyzed published studies of community health information outreach efforts. The goals were to identify characteristics of successful health information outreach programs, as well as effective evaluation practices for capturing best practices. In the area of consumer health informatics, we continued working on a study of factors that affect patients' comprehension of consumer health materials. Our FY2011 efforts involved working on two manuscripts detailing the findings of the study. The first manuscript was accepted for publication in the *Journal of Medical Internet Research*. The other, intended for submission to *Journal of Biomedical Informatics*, reached its final draft stage. The project was conducted in collaboration with colleagues from the University of Wisconsin, Madison, and Drexel University. Our third research area involves science education. Working with a collaborator from the teacher education program in the University of Maryland, College Park, we conducted a study of team interactions between teachers, educational researchers, and Web developers co-designing educational activities for environmental health education. The results of the study were accepted for presentation at the 2012 American Educational Research Association meeting. Finally, we initiated a study of the impact of college major and biological knowledge on college students' views on commonly held incorrect health beliefs. The study is a collaboration between SIS, City College of NY, and Georgetown University. The work conducted in FY2011 included study design and data collection.

# LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

*Clement J. McDonald, MD*  
*Director*

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

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

LHNCBC is organized into five major components: Cognitive Science Branch (CgSB), Communications Engineering Branch (CEB), Computer Science Branch (CSB), Audiovisual Program Development Branch (APDB), and the Office of High Performance Computing and Communications (OHPCC). An external Board of Scientific Counselors meets semi-annually to review LHNCBC's research projects and priorities. LHNCBC research activities can be found at <http://lhncbc.nlm.nih.gov/>.

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

These projects are efforts to target the overall recommendations of the NLM Long Range Plan Goal 3: *Integrated Biomedical, Clinical, and Public Health*

*Information Systems that Promote Scientific Discovery and Speed the Translation of Research into Practice.*

## *NLM Personal Health Record*

The goal of the NLM Personal Health Record (PHR) project is to help individuals manage health care for themselves and/or their relatives. The PHR is intended to serve as a test-bed for patient-specific and reminder-based consumer education information, 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 PHR supports the entry and tracking of key measurements, test results, prescriptions, problems, immunizations, and future health appointments. Patients can produce digital and paper copies of PHR contents in various formats. Patients can record questions they want to ask their doctors, maintain lists of their current medications, view educational material that pertains to their specific health information, and monitor trends in their weight, blood pressure or other items of interest. 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. These codes come from national vocabulary standards that are supported or developed by NLM, e.g., Logical Observation Identifiers Names and Codes (LOINC), RxNorm, and Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT). The strong use of coding in the NLM PHR vocabulary standards enables many computer-generated features such as personalized reminders about preventive care and/or healthy behaviors and automatic calculations based on other values in the PHR (for example, calculation of body mass index based on height and weight) and direct links to information sources such as MedlinePlus.

In FY2011, LHNCBC researchers and developers continued to improve the capabilities of the PHR including expanded and enhanced vocabulary and reminder rules, sophisticated graphing capabilities and auto-save features. Staff continued to work with NLM and others to develop and implement policies concerning the PHR. We worked with the NIH Office of General Counsel to write a software license agreement. We conducted an analysis of PHR user agreements and wrote a user agreement template. We began negotiations with a local hospital that may license the PHR software and provide NLM with de-identified data for usability research testing.

This young project addresses the long-standing 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

supports and will provide another consumer entry point to a rich trove of patient-oriented data.

#### *Use of Surescripts Prescription Data in Direct Patient Care*

Studies have shown that a significant proportion of Emergency Department (ED) visits are related to adverse events of drugs. It is vitally important that the ED physicians have access to a full and accurate medication history. However, gathering such information from the patient is time-consuming, expensive, and sometimes unfeasible when patients are unconscious. Patient-provided medication histories are also known to be incomplete. The ED of Suburban Hospital has employed the service of Surescripts, a consortium of major Pharmacy Benefit Managers (PBM) and the largest e-prescribing network in the US, to provide an electronic summary of a patient's full year prescription filling history. Before the system went live, Suburban Hospital collected both Surescripts data and patient-provided history for quality assurance. We obtained this information in a de-identified form. We compared the two sources of information and found that Surescripts information covered a high proportion (almost 90%) of a patient's current medication and added significantly to the manual history.

The Surescripts reports are now a routine information source for patients attending the Suburban ED. The nurses, pharmacists and physicians using the system have provided positive feedback about its use. In a high proportion of cases, drugs that are missing from the patient-provided medication report, but are present in the Surescripts report, are drugs that the patient is in fact taking. The full year prescription history is also helpful in identifying potential problems of drug compliance, drug seeking behavior, and duplicative prescriptions.

#### *EMR Database Research and Natural Language System Development*

Developed to investigate secondary use of data collected in electronic medical records, this general purpose longitudinal database structure proved to be robust when we added microbiology results obtained with the sixth update of the MIMIC II EMR dataset.

To facilitate the use of this database by researchers, in FY2011 we mapped the majority of the MIMIC-II local laboratory test codes to standard LOINC codes and incorporated them into MIMIC-II public releases starting with the sixth version. We are currently mapping medications from multiple tables to RxNorm nursing observations from MIMIC-II local codes to LOINC, RxNorm, and/or SNOMED CT as appropriate.

We also have used the MIMIC-II de-identified data under a restricted-use Memorandum of Understanding to conduct retrospective clinical studies on the significance of obesity and metabolic syndrome, glucose control, and vitamin levels in ICU mortality and post-discharge

survival. In the process of using the data for our research, we found problems that required correction and normalizing the clinical data within MIMIC-II.

We also use the MIMIC-II as a test bed for comparing speed and ease of use of a no-SQL database (SOLR) with a traditional SQL database (Oracle). For some kinds of queries, SOLR is 10-50 times faster than fully optimized Oracle. Because SOLR requires flattening of the database, it may provide an easier platform for the typical researcher to search: the researchers might no longer be burdened with understanding the relations between data elements and formulating complex SQL queries. We are developing an interface to test this hypothesis.

In line with the NLM mission to facilitate access to health information resources, this year we became a mirror site for PhysioNet, a very large database of physiologic wave form tracings gathered from health care institutions worldwide by the MIT researchers who also developed MIMIC-II. We have recently updated the MIMIC-II waveform data with the latest, 3-terabyte collection.

We developed natural language processing techniques to extract important clinical variables from the rich narrative text in MIMIC-II, e.g., smoking status and discharge destinations. For NLP experiments we acquired over 200,000 reports that were made available to participants of the 2011 Text REtrieval Conference (TREC). The UMLS lexicographers group uses these reports to augment the Specialist Lexicon with entries specific to clinical narrative. In addition, our research team participates in and regularly scores among the top finishers in the annual TREC Genomics track competition. The LHCBC team was this year's top-scoring team among 29 international participants from industry and academia.

#### **Biomedical Imaging, Multimedia, and 3D Imaging**

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

#### *Imaging Tools for Biomedical Research*

LHCBC and the American Society for Colposcopy and Cervical Pathology (ASCCP) launched one of our imaging systems, the Teaching Tool, for operational use in the assessment of professional knowledge and skills in the field of colposcopy. As of October 2011, the Teaching Tool is being used by 88 resident programs nationwide in Obstetrics/Gynecology and Family Practice (at 80 universities and other premier institutions such as the

Mayo Clinic). To date, the tool has been used to administer 760 individual online exams.

In addition, the National Cancer Institute (NCI) used another of our imaging programs, the Boundary Marking Tool, at the University of Oklahoma Health Sciences Center and other sites in Costa Rica, Nigeria, the Netherlands, and Spain to collect and annotate colposcopy images for the creation of a worldwide database for cervical cancer research. We incorporated new capability in our Multimedia Database Tool to retrieve and display histology images from the NCI ASCUS/LSIL (atypical squamous cells of undetermined significance / low-grade squamous intraepithelial lesion) Triage Study (ALTS), and we are currently working with NCI in a study of visual precursors of pre-cancer. This year, pathologists at the University of Oklahoma Health Sciences Center used the CEB Virtual Microscope to make visual diagnoses and graphically annotate histology images of the uterine cervix, and we are using this annotation data to develop computer-assisted diagnosis methods for these images. In the past year, this work has included research, design, and development of segmentation, feature extraction, and tissue classification algorithms for these images.

We also collaborated with academic researchers in projects to develop interactive segmentation capabilities for very large images using Graphical Processing Units (GPUs). Developers successfully installed this capability in an in-house system equipped with two GPU processors, and used it for the segmentation of Gigabyte-sized histology images. Additional collaboration with academic groups included work toward developing high-fidelity image compression techniques for mobile platforms and work in biomedical case-based (text and image) information retrieval.

### *Content Based Image Retrieval*

As a significant part of imaging research at LHNCBC, Content Based Image Retrieval (CBIR) is an active area with several objectives related to the development of tools and systems. One objective is to improve the state of the art in techniques to find visually similar images. A second goal is to extend image matching from simply finding images that are visually similar to those that are also meaningful in a particular context. This research helps introduce automation into our existing cancer research tools. For example, the 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. Evaluated in 2010 for usability and acceptance at the biannual meeting of the American Society for Colposcopy and Cervical Pathology (ASCCP), during FY2011 developers used the evaluation recommendations to improve the Boundary Marking Tool.

Investigators have used CBIR 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. Biomedical journal articles contain a variety of image types that can be broadly classified into two categories: regular images and graphical images. Graphical images can be further classified into four classes: diagrams, statistical figures, flow charts, and tables. Over 15 image features were implemented and used in a Support Vector Machine (SVM)-based framework to detect modality (x-ray, ultrasound, CT, MR, etc.) and to compute image similarity. Techniques developed in this work were evaluated in the international ImageCLEF competition since 2009 and were again found to be successful. In 2011 our efforts were ranked highly among 17 teams from around the world, many of which were from industrial research and development 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. Researchers continue to improve automatic image modality detection and annotation methods and are incorporating these findings into our OpenI search system.

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 individual figure panels from multi-panel images in biomedical articles. From each image panel author-placed markups (or “pointers”) such as arrows, asterisks, and alphanumeric characters are automatically detected. We developed novel methods that applied a combination of Markov Random Fields, Hidden Markov Models, Active Shape Models, and Particle Swarm Optimization techniques for each type of markup with over 90% accuracy in detecting arrows and alphanumeric characters. We also developed a neural network-based optical character recognizer that is specialized to recognize single characters in images, a challenging task for off-the-shelf commercial OCR packages. Further research is ongoing.

Investigators are also using CBIR in a new project for screening digital chest x-rays for pathology, such as tuberculosis and other pulmonary diseases, that are 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. Further 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 GPUs for compute-intensive CBIR tasks, with a particular focus on image segmentation. Through our collaboration with Texas Tech

University, we developed a method that uses GPU processing power for interactively following challenging object boundaries such as the separation between cancer and non-cancer cells in histology slides of the uterine cervix. We can then use the segmented regions to train classifiers to detect various stages of pre-cancer.

### *Interactive Publications*

The goal of this project is to conduct research into models for highly interactive multimedia documents that could transform the next generation of publishing in biomedicine. The project focuses on the standards, formats, authoring and reading tools necessary for the creation and use of such *interactive publications* that, in addition to text, contain media objects relevant to biomedical research and clinical practice: video, audio, bitmapped images, interactive tables and graphs, and clinical DICOM images such as x-rays, CT, MRI, and ultrasound.

In this project, LHCBC has created interactive publications containing these data types, developed tools for viewing and analyzing interactive publications (Panorama) and for authoring such documents (Forge). These tools are analogous to Adobe's Acrobat Reader and Professional for PDF documents. Panorama, written in Java, was one of 9 semi-finalists out of 70 entrants in Elsevier's Grand Challenge contest a year ago. Recent enhancements to Panorama include bar charts, and the capability to run natively on Mac OSX, following a formal usability study of the tool.

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

In order to avoid the need for large downloads which would be required by our current approach for some public, we investigated several Web-based approaches. In one approach, we modified Panorama to exploit Java's Web Start technology. This technology allows stand-alone Java software applications to be deployed over the network with a single click. In a second approach, we developed a Web browser version of Panorama (Panorama Lite) using Adobe Flex, thus eliminating the need to download the Panorama software. The only requirement to run it is to have Flash installed. Besides offering easy and intuitive usage, this client version has better line chart and graph support, and includes tables and subsets similar to the original Panorama. A feature unique to Panorama Lite is a Map View that can present data for example, at the county, state, and/or country level, in a color-coded form to readily visualize geographic patterns. LHCBC has recently

collaborated with a publisher to create two interactive publications from their full text articles. These are hosted on our Web site and available for use through either a browser (Panorama Lite) or with Web Start.

Considering the long-term development of Panorama and Forge in an open source environment, we have taken steps to support third-party developers to create, for example new viewers for electrocardiograms. For this purpose, we have ported the core code of these tools to support Eclipse plugins, enabling any open source developer to develop new functionality without needing to modify our code.

In a separate approach to new forms of publishing, LHCBC partnered with the Optical Society of America (OSA) to test the use of interactive publications in an operating journal. The goal was 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 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 independently from the publication.

Initial feedback from authors, editors, and readers was positive, though some readers faced a learning curve which included software reader program installation and navigation problems. We made mid-course corrections to solve most of these problems. The reader software remained radiology based and therefore posed a problem to readers who were not familiar with that interface. Eighty percent of readers reported an enhanced experience and 50% reported increased learning and understanding.

### *Screening of Chest X-rays for Tuberculosis in Rural Africa*

The LHCBC has begun a collaborative project with AMPATH (Academic Model Providing Access to Healthcare), an organization that implements the largest AIDS treatment program in the world, and supported by USAID. This new project exploits the convergence of imaging research and system development at LHCBC and NIH policy objectives in global health. Our objective is to leverage in-house expertise in image processing to clinically screen HIV-positive patients in rural Kenya for

lung disease with a special focus on tuberculosis (TB). Since chest radiography is important to the detection of TB and other pulmonary infections prevalent in HIV-positive patients, we have provided AMPATH with lightweight digital x-ray units readily transportable in rural areas. Their staff will take chest x-rays (CXR) of the population and screen them for the presence of disease. These x-ray units are already on site and are being readied for deployment by vehicle.

Since the lack of sufficient radiological services in the area suggests the need for automation to perform the screening, our in-house research effort focuses on developing software to automatically screen for disease in the CXR images. Our researchers are developing algorithms to automatically segment the lungs, detect and remove ribs, heart, aorta and other structures and then to detect texture features characteristic of abnormalities, leading to a 2-class discrimination between abnormal vs. normal case. These machine learning algorithms require sufficiently large training sets (i.e., example x-rays), and to acquire these many options were explored and IRB exemption received. We reached agreement with Montgomery County's TB Control Program which provided about 200 usable de-identified x-rays. We also received a small initial set of x-rays from the lightweight units in Kenya which are also suitable for training the algorithms.

Using these x-rays for training and testing, we are developing an algorithm for detecting lungs and ribs focusing on region-based features such as log Gabor wavelets that exploit the orientation of anatomical structures. A robust identification of lung shape plays a role in detecting TB in CXR since many abnormalities (e.g., pleural effusion) exhibit deformation in lung shape. After extracting the lung fields, the algorithm measures various geometric features that discriminate between normal and effused cases. Ongoing work is in identifying the most successful geometric features.

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

#### *Remote Virtual Dialog System (RVDS)*

The Remote Virtual Dialog System (RVDS) will make the NLM "Dialogues in Science" series, currently only available in the NLM Visitors Center, available anywhere through the Internet. Support for the project is coming from stimulus funds made possible through the 2009 American Recovery and Reinvestment Act. The project involves the enhancement of programmatic capabilities of the virtual dialogue model to make it sustainable and to

allow for expanded applications of the model. During FY2011, we developed a voice-to-text conversion and recognition tool which would be platform agnostic and introduce as little time delay as possible. Additions to the "Dialogues in Science" series and updates of some current series members are in the early stages of development.

#### *Computational Photography Project for Pill Identification*

Launched in September 2010, *Computational Photography Project for Pill Identification* (C3PI) is an authoritative, comprehensive, public digital image inventory of the nation's commercial prescription solid dose medications. This effort is directed toward content-based information retrieval (CBIR) to promote patient drug safety at the national level. Support for the project in FY2011 came from stimulus funds made possible through the 2009 American Recovery and Reinvestment Act.

Initially working in partnership with the NLM Specialized Information Services Division and the US Veterans Administration to study content-based retrieval methods for medical image databases, researchers developed computer vision approaches for the automatic segmentation, measurement, and analysis of solid-dose medications from these pilot datasets including work on robust color classification tools to help identify prescription drugs. Now, 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).

In FY2011, the contracted team generated over 25,000 images of 1,056 samples of solid-dose pharmaceuticals. The team generates high-resolution, high-quality reference images of each sample. In addition, the contractor is acquiring images from alternate cameras in a variety of lighting conditions to prepare a data collection for a broad effort or national CBIR challenge for the identification of medications. LHCBC staff are currently preparing a server-based repository and content management system to support distribution and curatorship of the collection.

#### *Virtual Microscope (VM) and Virtual Slides*

LHCBC 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 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 3,300 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 FY2011, staff continued to maintain two databases to record information about Visible Human Project use. The first, to log information about the license holders and record statements of their intended use of the images; and the second, to record information about the products the licensees are providing NLM in compliance with the Visible Human Dataset License Agreement.

### *3D Informatics*

The 3D Informatics Program (3DI) continued its research mission to address problems encountered in the world of three-dimensional and higher-dimensional, time-varying imaging. LHNCBC provides continuing support for image databases and continues to explore the growing need for image databases, including ongoing support for the National Online Volumetric Archive (NOVA), a collection of volume image data. This collection contains 3D data from across medicine. Contributors to the collection include the Mayo Clinic Biomedical Imaging Resource and the Walter Reed Army Medical Center Radiology Department. The archive contains such integrated and multimodal data as virtual colonoscopy matched with recorded video from endoscopic interventions, time-varying 3D cardiac motion, and 4D MRI of a human hand. In anticipation of new sources of data from research partners contributing to the Insight Toolkit, the 3DI group

is updating MIDAS software system and adding additional disk space. We are cultivating sources among research teams in confocal microscopy, and we are seeking collections derived from Visible Human Data including segmentations, annotations, and processed information. We continue to serve a broad community with these data, and seek to establish a leadership role through public data distribution.

Throughout FY2011, staff continued collaboration with the National Cancer Institute's Laboratory for Cell Biology and with teams within LHNCBC to visualize and analyze complex 3D volume data generated through dual beam (ion-abrasion electron microscopy) and cryo-electron tomography. This investigation centers on analysis of the spatial architecture of cell-cell contacts and distribution of HIV virions at immunological 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.

In FY2011, OHPCC has grown its commitment in high resolution electron microscopy research, expanding our work to include the processing of data collected through transmission electron tomography. We are currently attempting to adapt research software that uses graphics processing units (GPUs) for high performance computing for sub-volume averaging and reconstruction. We are working to develop the research implementation into mature production software for the study of protein structures on the surfaces of HIV and influenza virions.

In FY2011, the 3DI group continued to investigate the use of rapid prototyping technologies in Radiology. We analyzed the x-ray attenuation characteristics of the 3D-printing materials available at NIH. and are presently evaluating the use of contrast agents as printing materials to vary the appearance of the 3D models. In April 2011, we published our early work where we 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. Our ongoing effort now has begun to focus on resolution and contrast measurement of our methods to ensure the precision and accuracy of our radiological models. This work is conducted in partnership with the National Institute of Allergy and Infectious Diseases (NIAID).

## *Insight Tool Kit*

The Insight Toolkit (ITK) is a public, open-source algorithm library for the segmentation and registration of high-dimensional biomedical image data. 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.

This year, LHNCBC and the ITK Project coordinated the efforts of groups including General Electric Global Research, the Mayo Clinic, Harvard University, Kitware, Inc., CoSmo Software, the University of Iowa, the University of Pennsylvania, Ohio State University, Old Dominion University, Carnegie Mellon University, Georgetown University, the University of North Carolina at Chapel Hill, and the University of Utah Scientific Computing and Imaging Institute. The research topics supported by these software development efforts include microscopy, digital histology, tumor micro-environments, zebrafish embryology, deconvolution methods for astronomy and astrophysics, image

registration for neurosurgery, tumor volume measurement for lung cancer treatment, and video processing for security applications as well as healthcare. A beta release of ITK-version4.0 was released in September 2011 with a projected release of the stable version in December. Work is expected to continue until June 2012 with a planned release of ITK-v4.2. This work is funded through the American Reinvestment and Recovery Act.

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

This work seeks to exploit ongoing research in both natural language processing and content-based image retrieval by combining processing by text as well as by image features to index the open access journal literature. In this contribution to our Clinical Information Systems effort, techniques are developed to automatically identify relevant figures in biomedical articles (illustrations, clinical images, graphs, diagrams, etc.) that could provide multimedia assistance to clinical decision making. Following the recommendations of the September 2010 Board of Scientific Counselors that evaluated our initial prototype multimedia search engine, we focused on scaling up this system to large collections. In addition to redesigning the system architecture and refactoring the existing code, we improved the user interface, adding functionality (such as filtering images based on their type, filtering journals by clinical specialty, and ranking papers by clinical task, for example, treatment), and processing a larger set of scientific publications acquired from PubMed Central.

Our new experimental multimedia search engine, OpenI, retrieves and displays "enriched citations" - structured MEDLINE citations augmented by image-related text and concepts, and linked to images and image representations based on image features. The document and image processing system that generates these enriched citations, as well as the search engines, are hosted on a multicomputer cluster with a shared file system for distributing computing operations (Hadoop™ MapReduce) and enterprise-level bare-metal virtualization. Due to the demands of high-performance distributed computing, we implement virtualization that enables us to control the number of logical processors, as well as which logical processor runs on a specific physical core. This ensures predictable performance and scalability as well as tight resource control. Additionally, the cluster design implements the fault-tolerant features needed to ensure high availability of cluster resources. The system is capable of processing up to 65 concurrent information requests per second. The collection currently contains about 20,000 open access articles from PubMed Central, and 250,000 images. OpenI was demonstrated at various forums including the 2011 AMIA symposium, the University of Delaware, UC San Diego and elsewhere.

Staff also conducted research in key areas: ways to represent images with strings, these strings then indexed using traditional search engines such as Lucene; improved

methods to automatically segment multipaneled illustrations into single images, and to partition their captions to correspond to single images; improved methods to extract pointers (arrows, arrowheads, symbols) within images to identify regions of interest, among others. Steps have been taken toward building a visual ontology. We have also developed methods for segmenting lung and brain tissues and extracting key features for imaging properties of several pathologies in these tissues (for example, lung cysts, micronodules, and emphysema.)

### *Turning The Pages*

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

In FY2011 we developed an iPad application (app) containing two of these virtual books: Hanaoka Seishu’s *Surgical Casebook*, and Hieronymus Brunschwig’s *Liber de arte Distilland*. This app has been downloaded more than 3,300 times. We are currently adding Robert Hooke’s *Micrographia* to the iPad, and developing an iPhone version.

This year we released the kiosk, Web, and iPad versions for *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, and made in Mughal India.

We also launched a complete redesign of the TTP Web site, built around an open source content management system (Wordpress), while conducting research toward a realtime 3D version of TTP. Work in early 2012 will include adding two books to the kiosk and Web versions: Andrew Snape’s *The Anatomy of a Horse*, and a Mongolian prayer scroll.

## **Natural Language Processing and Text Mining**

### *Medical Article Records System*

In this project the goal is to introduce automation in many aspects of creating MEDLINE, especially in light of its rapid growth. The MEDLINE database now exceeds 20 million records. The Medical Article Records System (MARS) project, in operation for some years, aims to develop automated systems to extract bibliographic text from journal articles, in both paper as well as electronic forms. For the approximately 1000 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 FY2011, 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. Developers introduced three additional features to improve MARS performance to support: (1) expansion of the MEDLINE character set, (2) rezoning capability for Edit users, and (3) user interface for larger monitors used for the Reconcile (operator-verification) stage. These improvements required modifications to several subsystems, in particular the Edit and Reconcile modules.

Developers created and released a new system, Publisher Data Review (PDRS), in June 2011. This system is designed to provide data missing from the XML citations received from publishers: such as databank accession numbers, NIH grant numbers, grant support categories, Investigator Names, and Commented-on Article information. By providing these missing data, PDRS reduces the manual effort in completing the citations sent in by publishers, as well as correct their errors. The automated steps to fill in missing data and to correct wrong data substantially reduces the load on the operators, eliminating the need to look through an entire article to find this information, and then to key them in.

Investigator Names and Commented-on Article are the two most recent fields extracted by PDR. If done manually, extracting names of investigators is a particularly 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 open and read other related articles for commented-on information. We have designed and implemented machine-learning methods to extract these two fields.

To assist indexers to automatically retrieve “check tags”, indexing terms that are pre-defined (e.g., ‘human’, ‘female’, etc.), we developed the CheckTagger system. This prototype Web-based system locates such terms in the article being indexed, and presents them to the indexer for selection, eliminating the need to read through the article to identify the ter. The prototype was demonstrated to the Indexing section staff in October.

Staff are developing another system, WebMARS, to address cases where NLM is missing a journal issue or

when citation data from publishers is incomplete. WebMARS is a software tool that operators can use to automatically create missing citations from these problematic issues. This eliminates the current manual labor on part of the operators to 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 but this research also enables the creation of new initiatives in which these techniques find application, such as the ACORN project.

#### *Automatically Creating OLDMEDLINE Records for NLM*

The Automatically Creating OLDMEDLINE Records for NLM (ACORN) initiative aims to capture bibliographic records from pre-1960 printed indexes (e.g., IM, QCIM, QCICL, etc.) for inclusion in NLM's OLDMEDLINE database, thereby creating a complete record of citations to the biomedical literature since Index Medicus appeared in the late 19<sup>th</sup> century. This year 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 ACORN system consisting of three main components: Quality Control, Processing, and Reconcile. We have completed the Quality Control module and intend to deliver a pilot ACORN system in FY2012.

#### *Indexing Initiative*

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

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. Versions of MTI have also been created to assist in indexing NLM's History of Medicine book collection and for use in Cataloging. 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.

MTI's overall performance has dramatically improved this year due primarily to increased collaboration with Library Operations (LO) that has provided several rules based on their in-depth knowledge of the indexing process. Discussions with LO have also produced improvements due to a change in technical focus, emphasizing the precision of recommendations rather than recalling as many topics as possible.

Due to its success with certain journals, MTI was designated as the first line indexer for 23 journals totaling 3,211 articles. As a first line indexer (MTIFL), MTI indexing is still subject to the normal manual review process. The number of MTIFL journals will grow gradually and should prove to be a time and money saver for NLM.

The Indexing Initiative team also worked closely with two NLM Associate Fellows, whose projects were designed to explore automatic assistance to different aspects of the indexing process. One project evaluated the feasibility of automatically indexing comment articles for MEDLINE (30,000 each year). Results of the study showed that approximately 70% of terms assigned by indexers to comment articles matched terms assigned to the article being commented on. We added automation of the comment indexing to the DCMS system in October 2011. Automatically assigning the commented on article indexing to the comment articles will save approximately \$280,000 per year in contract indexing costs, while maintaining high quality indexing for these articles. The second project investigates the feasibility of automating the creation of functional annotations about genes, known as Gene Reference Into Function (geneRIF). This project is ongoing.

MetaMap is a critical component of the MTI system and a leading tool around the world in bioinformatics research. Recent work has provided significant speed improvement, XML (eXtensible Markup Language) output, negation identification, and user supplied acronyms/abbreviations list. MetaMap is also now available on Windows, Macintosh and Linux platforms. Users can now build their own data sets with the MetaMap Data File Builder and access their local version

of MetaMap via either an embedded Java API (Application Programming Interface) or UIMA (Unstructured Information Management Architecture) wrapper. We had approximately 700 downloads in 2011 for MetaMap, 180 for the Java API and 50 for the UIMA Wrapper. Of note, MetaMap is one of the NLM resources integrated in IBM's Watson system for healthcare applications.

### *Digital Preservation Research*

The Digital Preservation Research (DPR) project addresses an important mandate for libraries and archives: to retain electronic files for posterity as well as to retrieve information from preserved documents through semantic search. To preserve digitized documents, researchers have built and deployed a *System for Preservation of Electronic Resources* (SPER). SPER builds on open source systems and standards (e.g., DSpace, RDF) while incorporating in-house developed modules that implement key preservation functions: ingesting, automated metadata extraction and knowledge discovery.

NLM curators are using SPER to preserve more than 60,000 court documents from a historic medico-legal collection acquired from the FDA. In FY2011, NLM processed more than 20,000 documents and added them to a publicly accessible NLM Web site. In addition, SPER is being used to preserve another important collection, from NIAID, comprising conference proceedings of the "US-Japan Cooperative Medical Science Program on Cholera" (CMSP), a program conducted over a 50-year period from 1960 to 2010. Our activities toward this initiative include building a full repository for this collection with more than 10,000 documents, 2,500 research articles and names and affiliations of 6,000 investigators dealing with cholera. We extracted metadata from the document contents using automated metadata extraction (AME) techniques, and then built a portal for research articles, authors, investigators and institutions. The AME processes include: (a) layout analysis to recognize different types of information within a document set; (b) evaluating the effectiveness of models such as Support Vector Machine and Hidden Markov Model for different metadata layouts; and (c) capturing relationships among various entities in the collection from the extracted metadata.

Investigators are conducting research toward knowledge discovery from information preserved in this repository by (a) developing a domain-specific vocabulary, (b) generating RDF graphs or triples from the preserved information using this vocabulary and natural language processing techniques, and (c) building a knowledgebase accessible over the Web. The CMSP repository is being used as the prototype for this research task.

### *RIDeM/InfoBot*

As part of the Clinical Information Systems effort, the RIDeM (Repository for Informed Decision Making) project seeks to automatically find and extract the best

current knowledge in scientific publications. The knowledge is provided to several applications (OpenI – a multimodal literature retrieval engine, Interactive Publications, and InfoBot) through RESTful Web services. Developers expanded the services this year to extract salient information from patients' case descriptions.

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

The InfoBot API integrated with the NIH Clinical Center's EMR system, CRIS, is in daily use through the *Evidence-Based Practice* tab in CRIS since July 2009. Since then, we have found that the most followed links are to information about medications and protocols of clinical trials. These links are followed twice as often as links to definitions of terms, MedlinePlus and MEDLINE publications. Links to entry-pages of search systems (such as CINAHL) are used rarely. Investigators modified the API to accommodate changes to CRIS, without interruption of services.

### *De-identification Tools*

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

In a recent study, we tested four prominent state-of-the-art systems designed to recognize personal names in free text along with our Clinical Text De-identification (CTD) system. The CTD, with a hierarchical mean sensitivity greater than 99.9%, was better able to detect names in clinical reports than the methods of the other four systems.

We also developed a method to measure the risk of privacy breach by estimating the probability of identifying patients through the undetected personal names that remained uncensored in the text. In this study, CTD was the only system that fully protected the patient privacy without using hospital staff rosters and patient master index.

This study examines reports from one of the largest number of patient reported studies in the literature, using established as well as newly devised methods and metrics.

### *Librarian Infobutton Tailoring Environment (LITE)*

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

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

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

### *Terminology Research and Services*

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

The Lexical Systems Group recently began a project to enhance the derivational-variants function of the lexical tools. The derivational-variants function uses a set of derivational facts and rules to generate or identify derivational variants of input terms. Derivational variants are words related by a word-formation process like suffixation, prefixation or conversion (change of category). The current derivational variant system has only suffix rules and facts. These rules and facts are hand entered and curated. In order to add suffixation and conversion functionality to the system, the PDM team has developed a method to automatically extract candidate pairs of words that may be derivationally related, which helps automate the creation of rules and facts for suffixation and conversion.

The SPECIALIST Lexicon and Lexical tools are open source and freely downloadable. During the year, our Web page had an average of 3,229 unique visitors per month. We had an average of 1,700 downloads per month in 2011. We count 14 internal users, 27 known academic users and 22 known international organizations among our users.

The 2012 release of the SPECIALIST Lexicon will contain over 462,000 records, representing over 830,000 forms, an increase of over 13,000 records from the 2011 release. Many of the new terms are derived from de-identified clinical records from our own De-identification project and from the MIMIC database. We plan to further extend the lexicon by adding consumer-level medical vocabulary. The Consumer Health Vocabulary recently added to the UMLS Metathesaurus will supply some consumer terms. We also plan to obtain a frequency list of consumer terms quoted in clinical records through a collaboration with the University of Utah.

### *Medical Ontology Research*

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

During FY2011, staff investigated issues including quality assurance in ontologies, the representation of pharmacologic classes in biomedical terminologies, and approximate matching 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 LHCBC training program by providing mentorship to four graduate and two postdoctoral students, working with them on issues including data integration for pharmacogenomics studies, quality assurance in the UMLS, and Web services composition.

Research activities this year resulted in two journal articles, five papers in conference proceedings, two book chapters and five invited presentations. We continue to collaborate with leading ontology and terminology centers, including the National Center for Biomedical Ontology, the International Health Terminology Standards Development Organization (SNOMED CT) and the World Health Organization (ICD 11).

#### *Semantic Knowledge Representation*

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

The SKR project maintains a database of 60 million SemRep predications extracted from all MEDLINE citations. This database supports the Semantic MEDLINE Web application, which integrates PubMed searching, SemRep predications, automatic summarization, and data visualization. The application helps users manage the results of PubMed searches by outputting an informative graph with links to the original MEDLINE citations and by providing convenient access to additional relevant knowledge resources, such as Entrez Gene, the Genetics Home Reference, and UMLS Metathesaurus.

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

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

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

#### *ClinicalTrials.gov*

ClinicalTrials.gov provides the public with comprehensive information about interventional and observational clinical research studies. ClinicalTrials.gov receives over 50 million page views per month and hosts approximately 800,000 unique visitors per month. At the end of FY2011, the site had nearly 115,000 protocol records, nearly 4,700 of which display summary results, conducted in all 50 states and in over 175 countries. Approximately one-third of the studies are or will be open to recruitment, and the remaining two-thirds are closed to recruitment or completed. Data are submitted by over 9,300 study sponsors which include the US 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. Since that time, ClinicalTrials.gov has undergone enhancements to support other registration policies and to implement the requirements under Section 801 of the Food and Drug Administration Amendments Act of 2007 [Public Law 110-85]. In FY2011, 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 characteristics of the participants studied. Since the beginning of its operation, over 6,200 results records have been submitted by over 700 study sponsors. The average number of submissions per week has increased, with an average of 60 new results records submitted per week at the end of FY2011. The requirements for the expanded registry as well as the results database will be further elucidated through rulemaking. NLM is working with other institutes and centers and the Office of the Director at the NIH and the Food and Drug Administration (FDA) on a Notice of Proposed Rulemaking. The combined registry and results database provides access to critical information about ongoing and completed clinical research for patients, healthcare providers, and policy decision makers.

In FY2011, ClinicalTrials.gov was actively involved in educating the public and data providers on the new law, developing new features to improve the usability of the system, and promoting standards of transparency in clinical research through trial registration and results reporting. This information was communicated to a broad range of US and international stakeholders through presentations, workshops, a series of Webinars, 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 670 genetic conditions, more than 900 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 eighth anniversary in 2011. In the past year, the project expanded its genetics content for consumers. Specifically, GHR staff added more than 200 new summaries to the Web site in FY2011. We intend to continue this rate of production in FY2012, covering additional Mendelian genetic disorders as well as more complex disorders. The team also plans to continue expanding the gene families feature, which currently includes explanations of about 60 families of related genes. This year, the site averaged more than 21,700 visitors per day and about 33 million hits per month. GHR continues to be recognized as an important health resource.

This year, GHR staff performed outreach activities to increase public awareness of the Web site. 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 visiting journalists and students at a local university medical school. Staff members attended and represented the project at several major genetics conferences, and will continue to educate others about this useful resource in FY2012.

#### *Profiles in Science Digital Library*

The *Profiles in Science* Web site showcases digital reproductions of items selected from the personal manuscript collections of prominent biomedical researchers, medical practitioners, and those fostering science and health. *Profiles in Science* provides researchers, educators, and potential future scientists worldwide access to unique biomedical information previously accessible only to patrons able to make an in person visit to the institutions holding the physical

manuscript collections. *Profiles in Science* also serves as a tool to attract scientists to donate their collections to archives or repositories in order to preserve their papers for future generations. It decreases the need for handling the original materials by making available high quality digital surrogates of the items. Standardized, in-depth descriptions of each item make the materials widely accessible, even to individuals with disabilities. The growing *Profiles in Science* digital library provides ongoing opportunities for future experimentation in digitization, optical character recognition, handwriting recognition, automated image identification, item description, digital preservation, emerging standards, digital library tools, and search and retrieval.

The content of *Profiles in Science* is created in collaboration with the History of Medicine Division of NLM, which processes and stores the physical collections. Several collections have been donated to NLM and contain published and unpublished materials, including manuscripts, diaries, laboratory notebooks, correspondence, photographs, poems, drawings and audiovisual resources. The *Profiles in Science* collections are consistently popular. The Web site averages over 77,000 unique visitors each month. This year, the collections of pioneering surgeons Charles R. Drew and Clarence Dennis were added to *Profiles in Science*. Staff added 1,152 transcripts of documents, making handwritten items searchable and providing alternatives to PDF format files. Staff also added 31 digital items to the 33 existing *Profiles in Science* collections. Currently 141,583 image pages that constitute 26,868 digital items are available on *Profiles in Science*. The Web site now features the archives of 32 prominent scientists and health advocates.

The 1964–2000 Reports of the Surgeon General, the history of the Regional Medical Programs, and Visual Culture and Health Posters are also available on *Profiles in Science*.

#### *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 tablet devices, netbooks or portable laptops. PubMed for Handhelds (PubMedHh) requires no proprietary software and reformats the screen display as appropriate for the wireless handheld device being used. In support of evidence-based clinical practice, clinical filters feature easy access to relevant clinical literature. Newly developed resources allow searching MEDLINE through text-messaging. An algorithm to derive "the bottom line" (TBL) of published abstracts allows a clinician to quickly read summaries at the point of need. A "consensus abstracts" element provides rapid review of multiple publications with smartphones at the point-of-care. This corresponds well with a recent review of PubMedHh server logs that showed that more than 90% of queries were clinical in nature. Randomized controlled trials using

simulated clinical scenarios concluded recently at the Uniformed Services University, University of Botswana-University of Pennsylvania and the National Telehealth Center and Philippine General Hospital, Manila to evaluate the usefulness of abstracts in clinical decision making. We also developed and submitted an iOS (iPhone, iPad devices) app for PubMed for Handhelds.

### **Clinical Vocabulary Standards and Associated Tools**

Multiple projects in this area continue to promote the development, enhancement, and adoption of clinical vocabulary standards. The CORE Problem List Subset of SNOMED CT is published in the UMLS as a specific content view. RxTerms facilitates the use of RxNorm as an interface for medication orders. Inter-terminology mapping promotes the use of standard terminologies by creating maps to administrative terminologies, which allows re-use of encoded clinical data. The Newborn Screening Guide combines terminology and electronic messaging systems to facilitate care and research related to newborn screening. Another effort focuses on the development of a consumer-friendly medical problem and procedure terminology. LHCNCBC continues to play an important role in the UMLS project in research related to the various UMLS knowledge sources and providing support in UMLS production and user support. The inter-terminology maps are also available through the UMLS.

#### *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 (HHS). 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 US 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. We have an ongoing collaboration with the Mayo Clinic, Intermountain Healthcare, and Vanderbilt University to evaluate the CORE subset. Investigators completed a study of the use of post-coordination to expand the coverage of the CORE problem list. We released a new update of the Subset in August, based on the newest release of SNOMED CT and remapping of terms previously unmapped. The remapping process added 137 new concepts. The Veterans Administration has provided us with a new dataset which will add additional concepts to the CORE Subset.

#### *RxTerms*

RxTerms is a free, user-friendly and efficient drug interface terminology that links directly to RxNorm, the national terminology standard for clinical drugs. The Centers for Medicare and Medicaid Services has used RxTerms in one of their pilot projects in the post-acute care environment. It is also used in the NLM PHR. RxTerms is available for download from the NLM Web site. There is ongoing effort to align data elements between RxTerms and RxNorm. Investigators are currently reviewing the dose form information in RxTerms to improve usability. 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, SOAP-based and RESTful application programming interfaces (APIs) were created, 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.

During FY2011, staff released SOAP and RESTful APIs for two other drug information sources also 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. Usage of RxNav and the APIs has increased significantly, from 10 million queries last year to over 25 million queries in FY2011. Users include clinical and academic institutions, as well as pharmacy management companies, health insurance companies, EHR vendors, and drug information providers. In the future, an application facilitating the use of APIs will be developed, e.g., for mapping large amounts of terms and codes to RxNorm, for querying pharmacological classes (in NDF-RT) from codes in RxNorm, and for crosswalk purposes between drug vocabularies through RxNorm.

#### *Electronic Reporting of Units of Measure Standards*

During FY2011, LHCBC developed a Table of Example UCUM codes for Electronic Messaging, enumerating the Unified Code for Units of Measure (UCUM) units codes. UCUM codes are unambiguous and computable units of measure that include tables for converting one unit to another of the same dimension, e.g., pounds to kilograms. UCUM codes have been adopted as *the* unit of measure by many standards bodies. The example units table includes the 815 units of measure including the vast majority of units of measure used in clinical medical codes. This table will assist adoptees' use of UCUM in electronic health records, public health, and research projects. The content for this table was derived from Regenstrief Institute and Intermountain Healthcare.

We delivered this table to the HHS Office of the National Coordinator for Health Information Technology (ONC) Standards and Interoperability Framework Committee for their repository, and the Laboratory Results Interface Initiative for reference and inclusion in an HL7 Implementation Guide. Many organizations including HL7 and IEEE support UCUM, and groups in the US and Canada are already using the UCUM table, released in September 2011.

#### *LOINC Standards for Identifying Clinical Observations and Orders*

In FY2011, LHCBC continued to work with the Regenstrief Institute (RI), major laboratory companies, several NIH institutes, and other organizations to develop the size and breadth of the LOINC database. By the end of FY2011, LOINC had over 13,000 users in more than 140 countries and supported nine languages. We worked with RI and the LOINC Committee to create more than 6,000

new LOINC terms for both laboratory and clinical variables. These new terms included those from the PROMIS and PhenX survey instruments and were created as a result of collaborations with NIAMS and NHGRI, respectively. We continued to work 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.

We released the Top 2000+ list of lab tests that represent over 98% of the test volume carried by three large organizations that cover both inpatients and outpatients, as well as a companion Mapper's Guide that includes information on the development and organization of the Top 2000+ lab tests as well as advice and guidance about which codes to choose in specific situations (<http://loinc.org/downloads/usage/obs>). The goal of the Top 2000+ list and Mapper's Guide is to help hospitals, providers, laboratories, researchers, and others with the task of mapping their local codes to LOINC. We updated both the LOINC Web search tool (<http://search.loinc.org>) and the Regenstrief LOINC Mapping Assistant (RELMA®) so that the search results can be constrained to the top 2000+ tests.

During FY2011 we also reviewed, revised and distributed version 1.2 of the Common Lab Orders Value Set (<http://loinc.org/downloads/usage/orders>), which contains 332 laboratory tests that comprise over 95% of the lab test order volume in the US. We worked with RI to develop and publish the first version of The Table of Example UCUM Codes for Electronic Messaging (<http://loinc.org/usage/units>), which is based on content from Intermountain Healthcare and is a guide for mapping common lab result units to standard UCUM units. We also helped develop the HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-qualified Cytogenetics Model, Release 1.

#### *Newborn Screening Coding and Terminology Guide*

In collaboration with the Health Resources and Services Administration (HRSA), the Centers for Disease Control and Prevention (CDC), and the NIH Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), LHCBC has created and maintained LOINC codes for all variables used in newborn screening intended for delivery in an HL7 message implementation. Staff also invested special effort to develop a consensus approach to reporting findings of newborn screening for hemoglobinopathies, severe combined immunodeficiency (SCID), and lysosomal storage disorders (LSDs). These specifications are being adopted by many state newborn screening laboratories across the country. Kentucky, Oregon (which is the regional laboratory for five additional states), Colorado, New York, and Illinois are furthest along in the implementation work.

LHCBC is in the process of developing codes for newborn screening of Critical Congenital Heart

Disease in addition to the Recommended Uniform Screening Panel of the Secretary's Advisory Committee on Heritable Disorders in Newborns and Children. LHCNBC is also working with our partners to standardize data collection and coding for short- and long-term follow up, beginning with the laboratory tests for confirmation and diagnosis of conditions targeted by newborn screening.

### **Communication Infrastructure Research and Tools**

LHCNBC 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*

LHCNBC continues to investigate, review, and develop collaboration tools, research their application, and use the tools to support ongoing programs at the NLM. In our work with uncompressed high definition video over Internet Protocol (IP), we determined strengths and weaknesses of each of the three technologies (iHDTV, UltraGrid, and Conference XP) and we continue to overcome problems encountered in the delivery of uncompressed video due to differing platforms. We are monitoring the High Definition (HD) open source work of Video Conferencing Tool developers regarding H.264 compression. VIC is used by the AccessGrid, an open source collaboration tool widely deployed in universities and research centers and used in the OHPCC Collaboratory for research work and to support NLM programs. Finally, we are reviewing newer, cloud collaboration tools. We published a comparison of the major compression/decompression (codecs) available in the OHPCC Collaboratory for High Performance Computing and Communications (Collab), and we are in the process of writing systematic reviews of the uncompressed and cloud technologies the research team is studying. The overall research effort involves studying and testing collaboration technologies technically in our laboratory first and then deciding which warrant further applied research in clinical or educational settings and which might be put into use supporting NLM programs.

Until recently, iHDTV was the only uncompressed video system sufficiently robust to use in a clinical trial but OHPCC staff have worked with the developers of UltraGrid at Masaryk University in Brno, Czech Republic and it has improved considerably. Most notable has been the integration and synchronization of audio along with the video. We also provided video cards to the lead developer of ConferenceXP (CXP) at the University of Washington so he could directly test the programs he wrote to do uncompressed video for that system. The research team was only able to get the compressed video to work for CXP. The team concentrated on UltraGrid and ConferenceXP technologies because

development of iHDTV has ceased. Consequently, dual collaborations are planned with both Masaryk and the University of Washington, with a special interest in Masaryk, since the OHPCC and Masaryk research groups share an interest in 3D HD videoconferencing, 4K video, varied forms of HD compression, and the use of dynamic circuit networks (DCN) to ensure quality of service. The team will continue to collaborate with the Rochester Institute of Technology (RIT) to test open source software for compressed HD videoconferencing based on the H.264 video standard. RIT, Manchester University and others are refining the software and incorporating it into the AccessGrid, a technology used in the Collab for distance learning.

The installation of a 10 Gigabits per second (Gbps) network in the Collab has greatly facilitated collaboration with other institutions and our ability to test uncompressed video. Prior to installation, the technologies could only be tested back to back because they consumed bandwidth exceeding our network's capacity. The teams at NLM and Masaryk were able to conduct videoconferences using uncompressed HD video with synchronized audio over their research networks without the use of dynamic circuit networks (DCN) and to demonstrate this capability to the LHCNBC Board of Scientific Counselors. Eventually, DCN will be tested because traffic over the research network backbone is unpredictable.

Collaborators installed iHDTV systems at the Medical University of South Carolina (MUSC) to study uncompressed video's use as a diagnostic tool for teledermatology. We selected teledermatology as a research domain because previous research has shown it to be particularly difficult to use standard definition video to do remote dermatological exams. We measured diagnoses, clinician confidence, decisions to biopsy and physician and patient encounter satisfaction and compared under the following conditions: 1) when patients are examined in-person, 2) when patients are examined using uncompressed high definition video, 3) when patients are examined using compressed high definition video using a standard employed by all major commercial videoconferencing manufacturers, and 4) when patient data (history and photos) are used to assess patients by typical store and forward methods. iHDTV systems were chosen for the study because, at the time, audio was not integrated into UltraGrid and ConferenceXP could only transmit compressed video.

We completed follow-up research with MUSC on video medical interpretation. The follow-up study used lower quality video (less than full screen) and cell phone technology to assess video interpretation in a pharmacy setting at an outpatient clinic. We also did extensive tests with VSee, a low bandwidth video program. Results were mixed; however, the effort proved the feasibility of and requirements for doing video medical interpretation over cellular networks and we will publish the results in FY2012.

Staff continued to work with SIS on a distance education outreach program for minority high school students and with the NIH Library to offer NCBI database and other bioinformatics training at a distance. In FY2011, staff conducted bioinformatics programs with the University of North Carolina at Chapel Hill, the University of Tennessee at Memphis, and Virginia Commonwealth University.

#### *OHPCC Collaboratory for High Performance Computing and Communication*

LHNCBC established the OHPCC Collaboratory for High Performance Computing and Communication (Collab) as a resource for researching, testing, and demonstrating imaging, collaboration, communications and networking technologies related to NLM's Next Generation Network initiatives. Staff use this infrastructure to test new technologies of interest to NLM and to conduct ongoing imaging, collaboration and distance learning research both within LHNCBC and outside NLM. The facility can be configured to support a range of technologies, including 3D interactive imaging (with stereoscopic projection), the use of haptics for surgical planning and distance education, and interactive imaging and communications protocols applicable to telemedicine and distance education involving a range of interactive video and applications sharing tools. The latter enables staff to collaborate with others at a distance and, at the same time, demonstrates much of the internal and external work being done as part of the NLM Visible Human Project 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 last year to include compressed HDvideo and 3D display and DVD playback technology. This year we acquired a 3D camcorder for the purpose of using it and/or dual non-3D HD camcorders to transmit 3D HD video in future videoconferencing research.

#### *Computing Resources Projects*

The Computing Resources (CR) Team has a variety of core projects that builds, administers, supports, and maintains an integrated and secure infrastructure to facilitate the research and development (R&D) activities at LHNCBC and thereby augments the overall effectiveness of research projects. The integrated secure infrastructure contains network, security, and facility management, and system administration support for a large number of individual workstations and shared servers.

The network management includes the planning, implementation, testing, deployment and operation of high-speed networks over Internet and Internet-2. One core project implemented the 10-gigabit network, and studied many advanced communication protocols to support

LHNCBC collaboration activities and research projects. The network management team also participated in the study of Trusted Internet Connection (TIC) consolidation and evaluated the impacts to the NIH and NLM.

The security management team incorporates security operations into firewall administration, patch management, anti-virus management, intrusion monitoring, security and vulnerability scanning, and vulnerability remediation to ensure an IT working environment that is safe from overall security perspectives. One core project studied and implemented a unified patch management to improve LHNCBC's overall security measures. Another core project implemented an automated security audit system that ensures all systems at LHNCBC comply with policies. The security management team also studied and evaluated the network performance impact of Web anti-virus software to the NLM, and delivered secure ID to all servers.

The facility management team facilitates products' and servers' deployments, including power acquisition, network planning, cabling connection, and space allocation in the central computer room as well as at co-location facilities. One core project studied, designed, and implemented an enterprise console management system that enabled LHNCBC to remotely manage large numbers of servers.

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

#### *Disaster Information Management*

The Lost Person Finder (LPF) 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 and community-wide disaster scenarios.

The hospital-based LPF system includes means to photograph victims at the hospital's triage station, and to capture these pictures, general health status, and descriptive metadata (name, age range, gender, identifying features) using TriagePic, a Windows application for triage staff. This data enters a MySQL database which can be searched via a Web site or via Web services built by extensively customizing the open-source Sahana disaster

management system. The LPF system also features a "Notification Wall" that displays images of victims on both computers and 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. Because the LPF collects personally identifiable information (PII) such as pictures and names, we developed and received approval for a Certification and Accreditation Process and a System Security Plan.

One of the lessons learned in 2010 was the desirability of a unified Web site and database, capable of being used to respond to any disaster anywhere. We developed a prototype unified site, NLM Person Locator (PL) to hold data from multiple disasters, thereby eliminating the need to build multiple Web site and database instances. Staff further developed and tested the unified PL Web site during a multi-agency disaster drill (October 2011, Capital Shield 2012) at Suburban Hospital. We also deployed it for disasters worldwide, including the Christ Church Earthquake (February), the Japanese Earthquake and Tsunami (March), the Joplin Tornado (May) and the Eastern Turkey Earthquake (October).

For Capital Shield 2012, we ported our TriagePic application to a touchscreen tablet platform running Windows 7, and used by hospital staff successfully. The tablet has both forward (3 Megapixels) and rear facing cameras and a touchscreen interface allowing easy image and metadata capture by stylus or touch. Nurses and patient registrars at Suburban Hospital learned to use the tablet's camera and touchscreen interface easily, literally just before the drill started.

We continue to customize and enhance the Sahana open-source software, evolving suitable modules to address LPF's specific needs and findings from the drills and real disasters. Developers improved responsiveness through the development of SOAP-based Web service modules, enabling bidirectional communications of data to and from the Web site. We improved search capability through use of Solr/Lucene technology and SQL query optimization for low latency searching. Further, in an international exercise at a workshop in Spain, we tested automatic data interchange and mirroring with the Google Person Finder system in May using PFIF (Person Finder Interchange Format). We refined the automated mirroring process and used it in the Eastern Turkey Earthquake.

Ongoing research in this project includes:

- Developing techniques in image and face matching to de-duplicate records, detect and localize faces, and subsequently match faces using image feature-based methods.
- Experimenting with new communications protocols to collect victim data prior to arrival at a hospital, e.g., from ambulances.

- Refining Web services between TriagePic (or our iPhone app, ReUnite) and PL, to facilitate control and correction of reports sent to the site.
- Investigating the incorporation of our systems into a hospital's normal workflow, with a focus on the optimum mix of privacy, security and openness.
- Extending ReUnite to other platforms: iPad as well as the Android platform.

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

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

The NLM Media Assets Project provides the NLM with easy access to audio-video resources for improved biomedical communications. This includes:

- The Hypervideo Personal Digital Library/ Digital Video Library (a computer aided search, retrieval and viewing database).
- The NLM/History of Medicine Exhibits Audiovisual Assets Management.
- Archival management of the Visible Human Project film and digital image dataset.

The NLM Support Project provides NLM with the audio/video support and development needed to promote and augment NLM's operation. This includes:

- Support for the maintenance and operation of the NLM state-of-the-art auditorium, board room and conference rooms.
- Ongoing production, post-production, and authoring services for the development of Internet video, interactive multimedia for large-screen and tablet devices and displays, and Blu-Ray DVD production.

The LHCBC Research Support Project contributes to improving access to high quality biomedical imaging information. This project includes:

- The APDB/NCI collaboration on 3D visualization of molecular structures and functions in the discovery of disease and treatment.
- The Movement Disorders Database (a digital archive of movement disorders patients going through diagnostic routines).
- The Profiles in Science video modules.
- The Visible Human Project imaging and visualization research.

The LHCBC Core Resources Project provides research into developing new technologies for disseminating biomedical information. This project includes:

- The LHCBC Research Update Modules.
- Ultra High Definition Imaging Research.
- Ongoing design and development of image-rich Web sites in support of biocommunications.
- Audio/video/imaging archiving and asset management.

A number of LHCBC projects require videographics, interactive multimedia development,

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

Extensive work continued toward the planning and development of interactive multimedia for the FY2011 NLM Exhibition "Native Voices: Native People's Concepts of Health and Illness." APDB staff worked with the Director, NLM and CgSB staff to review the extensive video interview database to establish major content areas based on thematic indexing of all interview transcripts. Based on this work, APDB contributed to the successful production of interactive video kiosk programs within the exhibition which opened this year. We used the highest quality video format standards throughout, from production to encoding and compression and display. For this project we encoded, video in formats for distribution across multiple platforms including iPad and mobile smart phone applications, which are featured throughout the exhibition. Our focus on video compression codecs for small screen delivery, navigation, and search capabilities is an ongoing area of research related to the work of the exhibition as well as many other areas of NLM's information programs.

### **Digital Video Archive**

The extensive digital video library assembled for the NLM Director's exhibition interview database has added to APDB's ongoing effort in the digitization, organization and accessible storage of large-scale video libraries. Digital workflow management and file format standards established for exhibition production support are now being applied to a major project within APDB to convert the large, historical tape library within LHCBC, containing over four decades of NLM programs into a viable digital repository accessible for future use.

### *Biomolecular Visualization*

APDB staff continued to collaborate with the National Cancer Institute's Laboratory for Cell Biology and with OHPCC to visualize and analyze complex 3D volume data generated through dual beam (ion-abrasion electron microscopy) and cryo-electron tomography. In this work we focus on the analysis of the spatial architecture of cell-cell contacts and distribution of HIV virions at immunological synapses formed between mature dendritic cells and T cells.

APDB staff are working with NCI scientists on a novel imaging technique developed by the NCI: Ion-Abrasion Scanning Electron Microscopy, which slices and images stacks of data at the microscopic level. One recent exploration yielded a remarkable 3D volumetric model revealing the transfer of HIV from T-cell along filopodia to astrocyte. The final illustrative image was a finalist in

the National Science Foundation's Visualization Challenge.

The APDB-produced medical illustrations and animations have illuminated the character of several immunological cells, cell structures and their interaction with pathological viruses including HIV. 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.

### **Training and Education at LHCBC**

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

During FY2011, 61 participants from 15 states and seven countries received training and conducted research in a wide range of disciplines: 3D image processing, biomedical ontology research, biomedical terminology research, content-based information retrieval, de-identification of medical records, evidence-based medicine systems, image, text and document processing research, information retrieval research, literature-based discovery research, natural language processing research, personal health record research, pill identification research, research into collaboration tools, semantic Web research and systems for disaster management.

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

The ITP also sponsors a Clinical Informatics Postdoctoral Fellowship Program, funded by LHCBC, to attract young physicians to NIH to pursue research in informatics. This program is run jointly with the Clinical Center to bring postdoctoral fellows to labs throughout NIH. LHCBC continues to offer an NIH Clinical Elective in Medical Informatics for third and fourth year medical and dental students. The elective offers students the opportunity for independent research under the mentorship of expert NIH researchers. We also host a two-month NLM Rotation Program which provides trainees from NLM-funded Medical Informatics programs an opportunity to learn about NLM programs and current LHCBC research. The rotation includes a series of lectures showcasing research conducted at NLM and provides an opportunity for trainees to work closely with established scientists and fellows from other NLM-funded programs.

# NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

*David Lipman, MD*  
*Director*

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

NCBI is charged with providing access to public data and analysis tools for studying molecular biology information. Over the past 23 years, the ability to integrate vast amounts of complex and diverse biological information created the scientific discipline of bioinformatics. The flood of genomic data, most notably gene sequence and mapping information, has played a major role in the increased use of bioinformatics. Recently, clinical genotype and phenotype information 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 21<sup>st</sup> century.

NCBI supports a multidisciplinary staff of scientists, postdoctoral fellows, and computing professionals. NCBI staff 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 current genomics research to the development of public information resources.

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

In order to fulfill its mission, NCBI:

- Creates automated systems for storing and analyzing molecular biology and genetics information and associating it with related information in the biomedical literature.

- Performs research into advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules and compounds.
- Facilitates the use of databases and software by researchers and healthcare personnel.
- Coordinates efforts to gather and disseminate biotechnology information worldwide.

At the request of the NIH Director, during the spring and summer of 2011 an outside scientific public advisory group, the NCBI Needs-Assessment Panel, was convened to provide advice about NCBI support for the biomedical research community over the next three to five years, with particular emphasis on the needs of the molecular biology and genomics communities. This assessment was deemed critical given that NCBI is receiving a deluge of complex and diverse data resulting from advances in next-generation DNA sequencing technologies.

## 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 NCBI is computer/user support and biology research in genomic analysis.

### *GenBank*

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

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

GenBank is divided into separate divisions based on taxonomy and sequence data collection methods. Eleven taxonomy divisions (BCT, INV, MAM, PHG, PLN, PRI, ROD, SYN, UNA, VRL, VRT) contain

sequences for over 380,000 species. High-throughput sequencing divisions include GSS, ENV, STS, TSA and others. The Whole Genome Shotgun (WGS) division includes contigs (overlapping reads) from WGS projects. Annotations are permitted in WGS assemblies, and records are updated as sequencing progresses and new assemblies are computed.

The fastest growing division in GenBank is TSA (Transcriptome Shotgun Assembly), which contains shotgun assemblies of primary (mRNA) sequences deposited in dbEST, the Trace Archive, or the Short-Read Archive (SRA). Sequence records in the Patent Division of GenBank are collected from the US Patent and Trademark Office.

The amount of data submitted to GenBank continues to grow. GenBank's two major divisions, WGS and non-WGS, combined contain over 206 million sequence records and over 350 billion base pairs. The traditional nucleotide sequences division increased to 141 million records in FY2011 from 122 million in FY2010. The WGS division has grown to over 65 million records and 208 billion base pairs.

Substantial resources are devoted to the analysis and curation of sequence data. GenBank indexers with specialized training in molecular biology create the records, applying rigorous quality controls. NCBI taxonomists consult on organism classification, and, as a final step, senior NCBI scientists review the records for biological accuracy. GenBank now contains more than 1,200 complete genomes from bacteria and archaea, and 20 percent of these were deposited during the past year. The number of eukaryote genomes with significant coverage and assembly continues to increase as well, with over 460 WGS assemblies now available.

In order to simplify access to, and improve the quality of, the enormous amounts of data stored in GenBank, NCBI is continuously developing new tools and enhancing existing products and methods. Sequence data, both nucleotide and protein, are supplemented by pointers to abstracts and publishers' full-text documents as they become available. Links are provided to other 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 housed at the NCBI.

NCBI has developed various tools for GenBank data submission. Sequin is a stand-alone tool that updates and submits large groups of sequences to the database. Sequin version 11.75 was released in 2011. New features include submission wizards for viral sequences and ribosomal RNA/ITS/IGS sequences from cultured samples. The BankIt submission tool allows the author to submit sequences within an online form and validates submissions by flagging errors before the sequences are deposited.

## 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 RefSeq collection contains 13,137,813 proteins for 16,248 organisms, representing a 17 percent increase in the number of proteins and a 49 percent increase in the number of organisms over last year. There are a number of sub-branches for the RefSeq project with curatorial support provided by different sections for microbial genomes, viral genomes, plant and fungal genomes, higher eukaryotes and so on. Notably, the microbial section of RefSeq realized a significant amount of growth in the number of organisms (85 percent) due to the new RefSeq Targeted Locus project, which includes molecular markers used for phylogenetic analysis.

In the past year, over 33,000 new or updated RefSeq records were made public for the 41 taxa supported by the (primarily) transcript-based, curation-supported, process flow. This includes nine invertebrates, 29 vertebrates, and three plants. This count reflects the pool of RefSeq records that are considered well-supported and are available for updates by RefSeq curation staff (e.g., it excludes RefSeq model records produced by NCBI's eukaryotic genome annotation pipeline). RefSeq curation activities resulted in some aspect of content modification for 26,393 Gene records and review of 25,173 RefSeq records. In addition, the set of computationally added feature annotations was expanded. This included adding features for sequence motifs that are of biological interest including selenocysteine codons, upstream in-frame stop codons, mature peptide and other protein features (propagated from Swiss-Prot records), putative open reading frames that are considered to be noncoding due to nonsense-mediated decay, and signal peptide features as calculated by running an in-house copy of SignalPv4.0. In addition, to further support the goal of informing

customers of the availability and advantages of the RefSeq dataset, the NCBI Handbook chapter was extensively revised, and a RefSeq FAQ document was added to the NCBI bookshelf. In addition, the RefSeq curation group provides significant support to NCBI's whole genome annotation process flow by reviewing intermediate Quality Assurance (QA) results, providing QA testing for new data loaded to NCBI's genome browser (Map Viewer) prior to public release, and more. This support extends to the research community where some RefSeq curation staff serves as a point of contact for research communities to provide education about NCBI resources (particularly genomics resources), advocate for data submissions, and respond to community requests for curatorial improvements for transcript and protein representation for their organisms of interest.

The Consensus Coding Sequence (CCDS) database identifies a core set of consistently annotated, high quality human and mouse protein coding regions. Both human and mouse datasets were updated in the past year, yielding current totals of 26,473 CCDS IDs for human and 22,187 CCDS IDs for mouse. The total number of CCDS IDs grew by 17 percent. The CCDS project is supported by curatorial review so that annotation updates to proteins that have a CCDS ID are done in a synchronized manner that retains consistency in different international genome browsers. Updates are reviewed and must be agreed upon by members of an international collaboration. Public annotation is provided to explain the evidence for the curation decision to make a change. Over the last fiscal year, 494 entries were subject to collaborative review and 426 explanatory public notes were added to the database. In addition, NCBI hosted a small one-day meeting for CCDS collaborators and expanded the content reported to the CCDS FTP site.

The Gene database provides a unified query environment for genes defined by sequence and/or genes included in the Map Viewer. It integrates information about genes and gene features annotated in RefSeq and collaborating model organism databases. Gene is heavily used and manages information for more than 8.3 million genes from over 7,900 taxa. About 100,000 full records are retrieved each day from about 6,000 sessions.

This year much software development was focused on improving representation of phenotype information to support the upcoming Genetic Testing Registry (GTR) and other medical genetics projects. The Gene record display was also restructured to include an interactive display of genes on reference genomic sequences, including RefSeqGene, and to improve the reporting of genomic content.

In FY2011, the number of RefSeqGene records continued to grow, from 3,430 at the beginning of the fiscal year, to 4,405 at the end. Many RefSeqGenes were created in response to requests from Locus-Specific Databases, which need a stable reference standard sequence for each gene. RefSeqGene and LRG sequences

are also starting to be requested to provide sequence standards for practice guidelines in genetic testing.

Activity within the Locus Reference Genomic (LRG) collaboration also increased. Currently about 340 RefSeqGenes have been assigned LRG accessions, with 178 now public. As a corollary of establishing RefSeqGene, collaborations with the dbSNP and dbVar groups continued in order to establish sequence annotation of clinically important variants, and to release the information in a more timely fashion. Several suggestions about the human reference sequence were provided to the Genome Reference Consortium, based on the detailed analysis of genes being targets for RefSeqGene.

To support the medical genetics community for which RefSeqGene was implemented, key tools were developed to process human variation according to HGVS standards. These tools are a critical component of the infrastructure supporting a new tool to analyze human variation launched this year called Variation Reporter, which allows users to upload variants and receive a report of known variants and functional consequences.

Extensive testing, quality assurance, and documentation are essential to the release of data in Gene, Map Viewer, and BLAST, as well as documentation for Web sites that support the scientific community's access and use of NCBI resources. Gene staff is heavily involved in annotation of genes on genomic sequences via the genome annotation pipeline. In FY2011, improvement to the annotation pipeline process resulted in twice as many annotated genomes as any previous year. Twenty genomes were annotated for public use, thirteen of them for the first time. Some new genomes annotated include: *Meleagris gallopavo* (turkey), *Nomascus leucogenys* (Northern white-cheeked gibbon), *Bombus terrestris* (buff-tailed bumblebee), *Amphimedon queenslandica* (sponge), *Loxodonta africana* (African savanna elephant), *Arabidopsis lyrata* (lyrate rockcress), *Oreochromis niloticus* (Nile tilapia), and *Cavia porcellus* (Domestic guinea pig). Updates were provided to existing genome assemblies for 13 species including human (Build 37.2), *Bos taurus* (cow), *Rattus norvegicus* (rat), *Callithrix jacchus* (marmoset), *Macaca mulatta* (rhesus macaque), *Pongo abelii* (orangutan), *Ailuropoda melanoleuca* (giant panda), and *Apis mellifera* (honey bee).

The Map Viewer is NCBI's primary tool for visualization of assembled genomes. Genes or markers of interest are found by submitting a query against a whole genome or by querying one chromosome at a time. Cross-species comparison is supported by increased standardization of map features, and maps from outside sequencing centers are utilized for multiple-species queries. Query results are viewed in a results table that includes links to a chromosome graphical view where a gene or marker is seen in the context of additional data.

A new Genome Decoration Page (GDP) was released in FY2011 that allows users to place their own annotation data on a genome. Annotation can be shown as a track or set of tracks drawn next to the chromosome.

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 allows users to remap genome coordinates from one assembly to another. The tool uses the alignments of one assembly to the other to project annotation features. This year, the Remap service was updated to support mapping of assemblies from RefSeqGene sequences.

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, as well as generates the final assembly after collaboration and quality assurance. In FY2011, GRC released four updates of the human reference assembly (GRCh37).

UniGene is highly useful as a resource for cDNA clone selection and rudimentary expression profiling by many communities of scientists. Interactive usage of UniGene remains steady with 20,000 Web hits daily, 6,000 of these involving reviewing clusters. While UniGene has expanded to include 114 organisms, most of the 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 to remain relevant to users, most likely in the form of assemblies of transcribed sequences.

The HomoloGene database of homologous genes acts as a complement to Gene and UniGene. HomoloGene covers 20 model animal and plant genomes and provides a reliable and comprehensive database of gene homologs, offering statistics on inter-species sequence and protein domain conservation. The latest HomoloGene release, 65, includes updated annotations for human, zebrafish, fruit fly, *C. elegans*, and *A. thaliana*. Clusters also feature an updated, related UniGene section that groups linked UniGene records by organism.

The database of Short Genetic Variations (dbSNP) is a comprehensive catalog of common genetic variation. dbSNP contains over 178 million submissions of human genome data that has been processed and reduced to a non-redundant set of 52 million refSNP clusters. Over 100 other organisms (up from 84 one year ago) are represented in the SNP database, with 124 million submissions curated to 81 million refSNP clusters. In FY2011, new attributes added for searching and filtering of human SNPs include allele origin, clinical significance, global minor allele frequency, and suspect.

Usage of the UniSTS, a comprehensive database of sequence tagged sites (STSs) derived from STS-based maps and other experiments, remains steady, at about 2,500 hits per day. Work continues to integrate information in UniSTS with NCBI's Probe and dbSNP databases.

The Probe database stores molecular probe data together with information on success or failure of the probes in different experimental contexts. Nucleic acid probes are molecules that complement a specific gene transcript or DNA sequence and are useful in gene silencing, genome mapping, and genome variation analysis. The database contains over 11 million probes as of October 2011. The RNA interference (RNAi) resource stores the sequences of RNAi reagents and experimental results using those reagents, such as extent of gene silencing and a variety of phenotypic observations.

UniVec, NCBI's non-redundant database of vector sequences, has been updated to build 6.0. The UniVec database is used in conjunction with NCBI's VecScreen tool to screen nucleotide sequences for contamination with foreign sequences introduced during the cloning or sequencing process.

In FY2011 the number of sequences represented in the UniVec database increased by 12 percent. Newly added sequences include 167 complete vector sequences and nine adaptor, primer and multiple-cloning site sequences. The new additions include sequences that NCBI staff observed to contaminate many sequences submitted to GenBank, and also several sequences suggested by users. This update to the database will enable searches using the VecScreen tool to detect more of the foreign sequences added by various cloning procedures.

#### *Comparative Genome Data*

NCBI provides guides for comparing organisms on a genomic 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.

The 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. There are currently 1,813 eukaryotic genome sequencing projects completed, in draft form, or in progress.

The Genome interface was updated in FY2011 to the new Entrez standard. Search results and record views are discovery-oriented and feature the Discovery column with analysis tools and related data links. The site was completely redesigned this year and Phase I is now

available. Major improvements include a more natural organization at the organism level for prokaryotic, eukaryotic, and viral genomes. Also, reports include information about the availability of nuclear or prokaryotic primary genomes as well as organelles and plasmids.

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 2,973 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 2,546 genomes have been annotated in-house, and NCBI is working with 10 outside groups who submit data.

The Entrez Genome Project resource was replaced with the BioProject database. BioProject provides access to large-scale biomolecular projects such as single-cell organism genome, metagenome, and transcriptome sequencing, genotyping, variation and epigenomics. The resource redesign adds tracking of several data elements including more precise information about a project's scope, material, and objectives. A new browser has been introduced and organism overviews provide access to all projects and data for a particular species or strain. BioProject accession numbers have been added to each record, while Genome Project identifiers have been retained.

### *Specialized Databases and Tools*

NCBI's ClinVar database, which is scheduled to become operational in 2012, will facilitate access to and communication about the relationships asserted between human variation and observed health status. ClinVar will be collecting reports of variants found in patient samples, assertions made regarding their clinical significance, information about the reporter and other supporting data. The alleles described in the submissions will be mapped to reference sequences and reported according to HGVS standards.

NCBI is collaborating with NIH in development of the Genetic Testing Registry (GTR), which will provide an online resource for health professionals and the public to locate genetic test information that has been voluntarily submitted by test developers and manufacturers. The GTR will offer comprehensive listings of tests as well as

information about a test's scientific basis. The GTR is scheduled to be available in early 2012.

The database of genomic structural variation (dbVar) contains data on variant DNA less than or equal to one base pair in size. Submissions are accepted from whole genome comparative studies and locus and gene-specific data from quantitative studies. In FY2011, dbVar released newly designed Web pages that provide improved structure and organization. A total of 62 studies from seven organisms have been loaded into dbVar, for which data can be filtered by Study Type, Method, or Variant size. Variant call data are easily accessed such as available evidence of validation, clinical assertions, and allele origins.

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. Links are provided to other flu resources containing sequences, publications, and general flu virus information. Nearly 38,300 new influenza virus sequences were entered into NCBI's Influenza Sequence Database in FY2011. About 24,400 of them were from the NIAID Influenza Genome Sequencing Project, the NIAID-funded Centers of Excellence for Influenza Research and Surveillance, the Centers for Disease Control and Prevention, and about 60 other institutions worldwide, and were processed by the NCBI flu annotation pipeline.

The Flu Dataset Explorer provides an interactive tool for preliminary analysis of protein sequences from the Influenza Sequence Database or from a user's own file. New functionalities added to the Influenza Virus Resource include the ability to search sequences by WHO recommended vaccine strains, prototype viruses of well-defined lineages/clades and northern temperate, southern temperate, and tropical regions.

NCBI's 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. The Sequence Read Archive (SRA) archives data generated from massively parallel sequencing experiments. During FY2011, budget constraints necessitated a modification of focus for data collected in Trace Archive and SRA. The repositories now focus on high-throughput data that supports other types of data at the NCBI including: RNA-Seq, ChIP-Seq, and epigenomic data submitted to GEO; genomic and transcriptomic assemblies that are submitted to GenBank; and 16S ribosomal RNA data associated with metagenomics submitted to GenBank. SRA also is continuing to archive high-throughput data from a number of large studies funded by individual NIH Institutes.

The Clone Database integrates information about genomic clones and libraries, including sequence data, genomic position, and distributor information. Replacing the Clone Registry in FY2011, CloneDB is the new

resource for finding descriptions, sources, and detailed statistics on available genomic libraries for a large number of organisms. The new Library Browser allows filtering by organism, vector type, distributors, and a number of associated database end or insert sequences. The linked Clone Finder locates clones by chromosomal position or features or by regions bounded by two markers. Clone Finder is now available for human, mouse, rat, cow, horse, pig, and zebrafish.

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 contains over 785,000 records, an increase of 35 percent from last year. Approximately 17 billion new individual data points were accepted in FY2011, bringing the total number to over 50 billion (a 50 percent increase). The number of manuscripts that cite GEO accessions, or provide a direct reference to the database, increased by 40 percent this year, to over 14,000 manuscripts.

Existing GEO resource improvements were made this year as well. New repository browser and summary pages were developed for GEO data allowing users to more easily locate data related to their interests. Pipelines for handling next-generation sequence submissions and brokering to SRA databases were updated to further improve data capture. A new “Find Pathways” feature was developed allowing users to map thousands of differentially expressed GEO Profiles to a ranked list of pathways in NCBI’s BioSystems database.

NCBI’s BioSample database contains descriptions of biological source materials used in experimental assays. It has been populated with sample data directly from the primary data archives including SRA, dbGaP, EST and GSS. Currently, the database contains over 658,000 records. Basic submission requirements and formats have been developed for the database and will soon be available to the public. Collaborations are ongoing with ATCC, Coriell, and DMSZ repositories to have commercial samples represented in BioSamples.

Data from the NIH Roadmap Epigenomics Mapping Consortium deposited into GEO increased by 300 percent in FY2011. 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 contains 66 studies, 318 samples, and over 1,100 data tracks from five well-studied species. Data can be browsed by experiment or sample and a “Compare Samples” beta tool is available. Links were added to NCBI’s Epigenomics browser and to the new PheGenI resource, facilitating access to significant results from genome-wide association studies.

The NCBI Taxonomy Project provides a standard classification system used by the international nucleotide and protein sequence databases. NCBI’s rapidly growing Taxonomy database is curated to include the names of species for which sequences have been submitted to the protein and nucleotide databases. Over 28,500 taxa are represented in the database. The Taxonomy database browser can be used to view position in the taxonomic tree or retrieve data in any Entrez database for a particular organism or group. Searches may be made on the basis of whole, partial, or phonetically spelled organism names. The Taxonomy system also provides a “Common Tree” function that builds a tree for a selection of organisms or taxa.

The NCBI Genome Workbench is an integrated application for visualization and analysis of sequence data. It is designed to provide a flexible platform for development of new analytic and visualization techniques. Six releases were provided to the public in FY2011 and average monthly usage rose by more than 200 percent. In the same time period, the application was downloaded 7,300 times. The Web site for Genome Workbench was rebuilt to make content creation more accessible leading to more and richer tutorials, examples, and documentation. New features for displaying and working with phylogenetic trees were introduced along with many enhancements and bug fixes. The second half of the year saw Genome Workbench team’s focus transition from stability and support to customer engagement and feature enhancement.

Much of the development and visualization of Genome Workbench transfers directly to its Web-based counterpart, Sequence Viewer, which saw approximately 400 percent growth in usage this fiscal year. Sequence Viewer is designed to be an embeddable component to compliment other information-rich views. Another milestone for Sequence Viewer was the construction of a browser for external curators to study data collected on children with developmental issues. There are at least six NCBI users embedding Sequence Viewer for onsite use, including Elsevier. Several Genome Workbench tutorial videos were added to the NCBI YouTube channel.

### *Chemical Information*

PubChem is organized as three linked databases: PubChem Substance, PubChem Compound, and PubChem BioAssay. Together, they form a complete information resource for millions of small molecules, including their bioactivity data, structures, and properties. The PubChem databases are a key component in the Molecular Libraries and Imaging initiative of the NIH Roadmap. Over 85 million substances have been deposited into PubChem over the past seven years, representing over 30 million chemically unique compounds.

PubChem BioAssay is the public repository of bioassay results for small molecules and RNAi. Users can examine descriptions of each assay’s parameters and

readouts, with links to substances and compounds. The number of bioassay records increased significantly in FY2011 and the BioAssay data model was further expanded to accommodate data diversity. Bioassay records reached 500,000, representing over 300 million experimental bioactivity test results.

During FY2011, BioAssay presentation in Entrez was updated to the most recent Entrez style. New Entrez index fields were added to improve data discovery and several new data analysis tools were released to support bioactivity analysis and data downloads. New Target and Compound/Substance views for the BioActivity Analysis service allow users to explore bioactivity from molecular target, chemical structure, and biological experiment perspectives. Finally, the bioassay deposition system was improved to provide instant feedback to depositors through the newly designed Preview tool.

The PubChem Substance database contains chemical substance records and associated information. The PubChem Compound database provides unique chemical structures and validated chemical depiction information describing substances in the PubChem Substance database. The naming of PubChem Compound records was improved by eliminating synonyms that were not used in a structurally consistent way to name chemicals and also by a new voting algorithm to better match MeSH to PubChem records by using structurally consistent synonyms. The Compound/Substance summary page was updated to include substantially enhanced annotation per chemical structure along with an updated look and feel. The increased annotation includes categorization/classification trees for drugs and small molecules from KEGG, ChEBI, and DrugBank.

PubChem 3D was released with major updates in FY2011. Changes include the addition of multiple conformers per compound with a diverse ordering to give maximal diversity with a limited set of conformers. Also added is the ability to analyze chemical activity by means of 3D similarity via single linkage clustering and heatmap displays. A precomputed 3D similarity relationship for all small molecules was added as well as the ability to search chemicals by 3D similarity and visualize 3D similarity relationships. Finally, users can now compute 3D similarities for arbitrary sets of compounds and conformers in PubChem and download 3D structures and 3D images.

#### *Protein Information*

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

#### *Molecular Modeling Database (MMDB)*

NCBI's Molecular Modeling DataBase (MMDB) 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.

In May 2011, a new version of Cn3D (V 4.3) was released to the public. New features include the ability to visualize the content of biologically relevant structural units, as defined by the authors and providers of 3D structure data records. Biologically relevant units may include molecules generated by applying transformations from crystallographic symmetry, which are now pre-computed and provided by MMDB. Cn3D 4.3 comes with side-by-side stereo views, additional alignment algorithms for editing multiple sequence alignments, and new highlighting features. Similar to the Cn3D 4.2 preview release that was distributed earlier as a bundle with CDTree, it has been packaged as a standalone program and enhanced to handle the new MMDB data specification that now includes biological units and interactions.

New MMDB structure summary pages featuring biological units and interactions were presented to the public. The summary pages have been revised to display salient features of each structure, including its biologically relevant structural unit(s) and an interaction schematic depicting the interactions among the structure's molecular components. The graphical summary is augmented by a tabular listing of molecular components and their interactions, which provides functional annotation, as well as links to structure neighboring and other resources. The MMDB help documentation has been updated accordingly.

In July 2011, NCBI's Structure database was presented with a revised home page, search interface, and search results display, to mirror the new Entrez design. Changes include: a streamlined home page with links to related resources; an "Advanced Search" page, which provides the ability to build a query one term at a time, browse the index of any search field, and combine earlier searches; and new search results displays that provide links to search filters, related data, and tools.

In FY2011, much of the groundwork was laid for further enhancement of structure services: the handling of "merged" structure records (as the Protein Data Bank splits up some experimental data sets into separate records, due to limitations of their data distribution format), and the computation of structure neighbor relationships of biologically relevant complexes, which may involve more than one protein molecule.

The Entrez 3D Domains database was shut down and removed from Entrez. 3D domains are still used in structure neighboring by VAST, but need not be represented separately in a structure neighboring approach that focuses on biologically relevant structural complexes.

The interaction tracking database and a corresponding Web service termed IBIS (Inferred Biomolecular Interactions Server) have 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.

#### *Conserved Domain Database (CDD)*

At the end of FY2011 a new version of the Conserved Domain Architecture Retrieval Tool (CDART) was released that offers new functions and features for finding proteins that have domain architectures similar to a query protein. CDART has been reengineered for improved performance with long query proteins and complex domain architectures. An expanded selection of input options now allows users to enter queries as protein sequences, sets of conserved domains, or multiple queries at once. CDART now displays simple similarity scores to make the ranking of hits more transparent. New options to filter results add the ability to include/exclude proteins from specific organisms and/or proteins that contain specific conserved domains, and to combine the filters with Boolean logic.

In July 2011, the Conserved Domain Database (CDD) was updated to the new Entrez standard. Changes include: a streamlined home page with links to related resources; an "Advanced Search" page; and new search results displays. Eight releases of CDD were processed and published in FY2011, going from version 2.25 to 2.32. These included a mirror of the most recent SMART collection of domain models, a mirror of Pfam version 25, TIGRFAM release 10, and the fall 2010 release of the Protein Clusters database. 1,943 new or updated NCBI-curated models were released. CDD curators checked in 3,177 unique models as curated into the tracking database.

The Batch version of CD-Search was released at the beginning of FY2011, with usage increasing slightly throughout the year. The batch version of CD-Search supplements the widely used CD-Search Web application and can process up to 100,000 queries at a time; it returns results in various formats to support easy parsing and interpretation. It can be used interactively or utilized from within computer programs or scripts to support pipelines for the large-scale annotation of protein sequences, such as in the analysis of whole genomes.

#### *NCBI Biosystems Resource*

In September 2011, NCBI's BioSystems database was updated to the new Entrez standard. Since August 2011, the BioSystems database includes additional records from KEGG, including pathway modules, which represent subsets of larger pathways, as well as structural complexes and functional sets.

As of March 2011, Gene expression data in the GEO Profiles database link to associated records in the BioSystems database. A button on the GEO Profiles search results page will automatically perform a multiple step "FLINK" operation that retrieves pathways containing one or more of the genes represented in the expression data. The pathways can be ranked by frequency or percent coverage. Also, WikiPathways data are now available in the BioSystems database. WikiPathways is an open, public platform dedicated to the curation of biological pathways by and for the scientific community. In addition, the BioSystems database continues to be updated weekly, as new data become available from the source databases. Records in the Gene database now link to corresponding pathways in the NCBI BioSystems database and allow users to view the gene within the context of the pathway diagram.

#### *BLAST Suite of Sequence Comparison Programs*

Comparison, whether of morphological features or protein and DNA sequences, lies at the heart of biology. BLAST has made it possible to rapidly scan huge sequence databases for similar sequences and to 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 and expanded for effectiveness and ease of use. BLAST now has 15 specialized pages corresponding to NCBI databases. The dbSNP BLAST page was updated in FY2011 with an improved submission form and output format. The new pages also have improved organism selection and chromosome-specific database selection, as well as other features of BLAST. Genome-specific BLAST pages now use the standard BLAST form, allowing users to have the full functionality of BLAST, such as adjusting algorithm parameters while searching a specific genome. Transcript Shotgun Assembly (TSA) BLAST database was made available this year as well. A new SOAP-based BLAST service was added to the BLAST suite. This service makes use of the Simple Object Access Protocol (SOAP) to submit and retrieve searches with the BLAST Web server.

Stand-alone BLAST version 2.2.5 was released this year as well. Improvements include hard-masking of databases, faster formatting of databases using makeblastdb, XML and best hit options for Blast2Sequences, multiple query psiblast, selection of any master sequence in psiblast with multiple alignment input, and query and subject length in tabular output. Detailed set-up instructions for standalone BLAST are now part of the BLAST User Manual on the NCBI Bookshelf.

## **Integration of Clinical, Genetic, and Environmental Databases**

The database of Genotypes and Phenotypes (dbGaP) was originally built to house, display and distribute data produced in Genome Wide Association Studies (GWAS). However, the mission of dbGaP has expanded beyond GWAS to represent evolving technologies and data types. The data that make up dbGaP studies can typically be categorized into three main classes: Phenotype data consisting of clinical, anthropomorphic, demographic and exposure variables collected from a few hundred to thousands of individuals; Molecular data produced using SNP chip arrays, expression arrays, epigenetic assays and/or sequencing of RNA or DNA; and Study Documents consisting of protocols or data collection instruments that describe how data or samples were collected and/or processed. The dbGaP system has proven that it is readily adaptable to house data from many different types of studies that include human phenotype information.

### *Study Submissions*

dbGaP studies are comprised of public summary-level data and/or individual level phenotype/genotype data and/or high throughput sequencing data. The database has released 177 studies since 2007, with over 66 studies released since last year. Notable studies released in FY2011 include five studies from the NHLBI-funded Grand Opportunities-Exon Sequencing Project (GO-ESP) of 800 sequenced exomes.

### *Linking Studies to Other NCBI Resources*

All dbGaP studies are cross referenced in the new BioProjects database at the study admin and data levels and links to funding BioProjects will be a focused effort for dbGaP staff in the coming months. Currently, disease traits for each dbGaP study are linked to appropriate Medical Subject Headings (MeSH) terms, as created by the submitting investigator. Samples for each study in dbGaP contain links to the Entrez BioSample database. In addition, links to PubMed articles are included for each dbGaP study.

### *Cumulative Summary Counts*

Collectively, dbGaP studies released to date consist of measurements for 296,758 total research participants, and include:

- Over 128,743 individual level traits (dbGaP variables).
- 2,770 documents linked to 150,392 variable summaries. Documents describe the studies through natural language descriptions, collection forms, or scientific protocols, and provide models for future research. Links between searchable documents and

variable descriptions provide an unprecedented level of functionality and usability.

- There are now trillions of individual genotypes (single-nucleotide measurements of participant DNA sequence) that potentially reveal systematic and heritable genetic differences between affected and unaffected individuals.
- 3,110 pre-computed statistical associations describe locations in the human genome where differences between affected and unaffected participants are statistically significant.

### *Authorized Access System Download Activity*

Principal Investigators (PI) request access for individual-level data submitted to dbGaP via the Authorized Access System. As of December 2011, 1,984 PI-defined research projects have been created in the approval system. Each description of a proposed research activity contains context and justification for 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 Committees (DAC) for review. DAC review confirms that proposed research use is consistent with the data use limitations placed on the data by study participants during the informed consent process. Approved users return to the dbGaP system to create and download a password-protected copy of the de-identified individual-level data (i.e., phenotypes and genotypes) to their local secure computing environment.

### *Data Usability: Tools and Software Development*

During 2011, 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. GaP Plus has also been incorporated in the Phenotype-Genotype Integrator (PheGenI), a user-friendly Web interface that integrates various NCBI genomic databases with association data from the NHGRI GWAS Catalog and supports Web browsing and downloads of published GWAS 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 research. The ability to traverse the

literature and the molecular sequences via “neighbors” and links provides an efficient and intuitive way of accessing data. Entrez currently supports and integrates 38 databases, providing sequence, taxonomy, gene, chemical, and biomedical literature and data.

### 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 formerly unknown 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, Entrez databases other than PubMed began incorporating sensors and other discovery-related features into their search results pages.

The PopSet pages were redesigned in FY2011 and fully integrated into the updated Entrez system. PopSet is NCBI’s database of related sequences and alignments from phylogenetic, population, mutation and ecosystem studies. New features include embedded graphical alignment and better integration of related data from other PopSets and Entrez databases.

NCBI’s Bookshelf updated its pages to the discovery-oriented design of Entrez including a new search results display, new limits features, and new advanced search. New and updated titles are highlighted on the new page, as well as a new “Featured Titles” section, with books picked by the Bookshelf team.

The PubChem databases incorporated their site function to include the recent Entrez design. PubChem BioAssay added new Entrez index fields in order to enhance data discovery. In addition, PubChem Compound improved its voting algorithm to better match MeSH to PubChem records. Another example of improved discovery potential is updated links from Entrez Gene to BioSystems that provide a gene view within the context of the gene’s biosystem pathway diagram.

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

PubMed is continually updated and enhanced for better functionality and more precise search results. In FY2011, the PubMed Limits page was modified to include two new subsets: Dietary Supplements and Veterinary Science. The Journals database was retired, merging with the NLM Catalog. Currently, the NLM Catalog contains all journal information. The MeSH database was updated to reflect the most recent Entrez interface design.

In March of 2011, PubMed Mobile Beta was released. This special Web interface makes PubMed faster to load and easier to use on mobile devices. It includes the same basic search functionality and content as Standard PubMed, without advanced search functions such as Limits, My NCBI, or Link/Outside Tool. PubMed Mobile replaced the PubMed Text version that was created for those requiring assistive technology.

My NCBI is an Entrez feature that allows users to store searches and results, with the option of automatically updating searches and sending results via e-mail. A new My NCBI interface was unveiled in FY2011. With the new release, all functions are viewed directly from the My NCBI homepage, where they are readily accessible for set up and customization. Searches can be done directly on the homepage with a “Search NCBI Databases” link along with a “NCBI Site Preferences” link that allows users to set up display preferences. A new Collections feature allows results and records from any NCBI database to be saved in My NCBI. The My Bibliography feature was updated to include more features for eRA Commons users as well as a Public Access Compliance wizard. A BLAST History tool was added as a custom option. A tutorial video was added to the NCBI You Tube channel introducing new features and updates to My NCBI.

### LinkOut

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

In FY2011, the number of organizations participating in LinkOut increased to over 3,400, representing an eleven percent growth rate over the past year. LinkOut providers include 2,664 libraries, over 440 full-text providers, and 295 providers of non-bibliographic resources, such as biological and chemical databases. Participation in Outside Tool, a service linking to external resources, increased to over 850 institutions. LinkOut users can now link to 87 million Entrez records, including links to the full text of 59 percent of PubMed records from over 9,300 journals. Usage of LinkOut resources reached over 35 million hits per month, and about 1.35 million hits per weekday.

A new LinkOut indexing process was put into place in FY2011. The redevelopment enhanced the support for UniCode and non-English languages, and improved efficiency and reliability of LinkOut indexing. A number of improvements were made to the Library Submission Utility. A CSV option was added as a means to obtain LinkOut statistics. The LinkOut file validator was improved to provide feedback related to additional error situations. A new utility was added to provide link resolver vendors with easy access to their clients' LinkOut ID and Name, which are used to submit LinkOut files on their behalf.

A utility for full text and non-bibliographic providers is in the planning stage. The goals are to easily correct errors due to redundant submission, allow providers to send links using a utility rather than XML files, and to provide a clear picture of existing link patterns for better quality control.

#### *PubMed Central*

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

PMC now has more than 1,200 participating journals that deposit either all their articles, or at least all the NIH-funded articles that they publish. For publication year 2010, PMC received over 70,000 NIH-funded articles and an additional 100,000 peer-reviewed papers from participating publishers.

Use of PMC continues to increase in concert with the growth of available articles. On a typical weekday, PMC has up to 500,000 unique users retrieving over 850,000 articles. As of October 2011, there were just under 2.3 million articles in PMC.

#### *Bookshelf*

The NCBI Bookshelf gives users access to the full text of over 600 textbooks and documents in life sciences, medicine, and healthcare. In addition to textbooks from commercial publishers, the Bookshelf includes tutorials and help documents authored by NCBI, NLM, and NIH staff. In FY2011, the Bookshelf updated its Webpages

with a new homepage, new displays of search results, new limits features, and a new advanced search. Highlights include a Browse Titles link where users can peruse a list of titles related to a term or subject of interest. Also, an option of viewing results by relevance or book has been added. All book displays have been updated in order to improve navigation.

The Bookshelf introduced the concept of the NLM Literature Archive (NLM LitArch), to make it easier to share book content in relevant Web sites. Currently, content in the books database is being displayed in Bookshelf and in PubMed Health. In the past, there was no distinction between the name of the database and the access service, which was Bookshelf. This year, 296 titles were added to NLM LitArch that include technical reports, books, documentation, and databases.

The OMIM database is now an independent site at omim.org while NCBI databases continue to link to OMIM records. NCBI continues to host the GeneTests and GeneReviews resources, developed at the University of Washington. Together, the resources support integration of information on genetic testing and disorders to facilitate the care of patients and families with inherited conditions.

#### *E-Utilities*

Entrez Programming Utilities (E-Utilities) are a set of server side programs that use a fixed URL syntax that translates a standard set of input parameters into the values necessary for software components to search for and retrieve data. All E-Utility programs are described in the Entrez Programming Utilities Help Manual.

EFetch is one type of Entrez Utility that retrieves records by unique identifier. EFetch was updated in FY2011 to include the BioSample, BioSystems, and SRA databases as well as defined default values for retrieval mode and retrieval type.

#### **Research**

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

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

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

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

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

The Board of Scientific Counselors (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 36<sup>th</sup> meeting was held in April 2011.

### **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 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 information to various databases, and obtain information about NCBI products and services.

Support for users of NCBI's services continues to grow in conjunction with the exponentially increasing number of submissions, updates, and usage, as well as increasing complexity of data. 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.

To keep up with ever increasing requirements for computing power and storage capacity, NCBI added over 3,000 CPU cores to its network and over 1 PB of disk storage in FY2011. The additional CPUs went largely to NCBI's grid-based "compute farm" and to its BLAST service. The storage was required by the SRA and related genome-scale sequencing projects. Also in support of SRA, 3 PB of tape-based data archiving capacity was added both to NCBI's primary Bethesda data center and to a colocation facility in Sterling, Virginia.

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. This team designed and put into use a content management system for NCBI pages in FY2011.

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.

### *Outreach and Education*

NCBI's 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.

NCBI's social media presence increased throughout FY2011. Usage of NCBI's Facebook, Twitter, and YouTube sites increased markedly. NCBI added numerous tutorials to its YouTube channels including:

- Save Searches and Set E-mail Alerts
- Sequence Viewer: Using Primer-BLAST and Using BLAST
- Sequence Viewer: Six Frame Translations
- Genome Workbench: Six Frame Translations
- PMC 10<sup>th</sup> Anniversary Video
- BLAST Results: Expect Values Part 1 and 2
- Assign Downloaders for dbGaP Data
- PhnGenI: Introduction to the Phenotype-Genotype Integrator

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

### *Training*

The NCBI Education Program provided Discovery Workshops and Webinars aimed at various types of users, from beginners to seasoned researchers. Ten Webinars were conducted to six institutions in FY2011. Webinar topics include updates to BLAST, updates to genomes, and NCBI overviews.

NCBI conducted three "Discovery Workshops" in FY2011. Workshops were held on the NIH campus and at Washington University, and taught by NCBI service desk staff. The course consists of a lecture and 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.

A seminar was provided at the American Society of Human Genetics (ASHG) annual conference for biomedical researchers and genetic professionals. The session was presented by various NCBI staff members, demonstrating NCBI Web and client-side applications used to obtain, manage, and analyze genome-scale data. Topics included: Getting the Most from the Reference Human Genome Assembly; Variation Resources; Genotype and Phenotype Information; and Medical Genetics Resources.

# EXTRAMURAL PROGRAMS

Valerie Florance, PhD  
Associate Director

The Extramural Programs Division (EP) administers extramural grant programs for NLM as authorized by the Medical Library Assistance Act (MLAA) and Public Health Law 301. EP's first grant awards were issued in 1965. The funds are expended as grants-in-aid to the extramural community in support of the Library's mission. Review and award procedures conform to NIH policies.

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

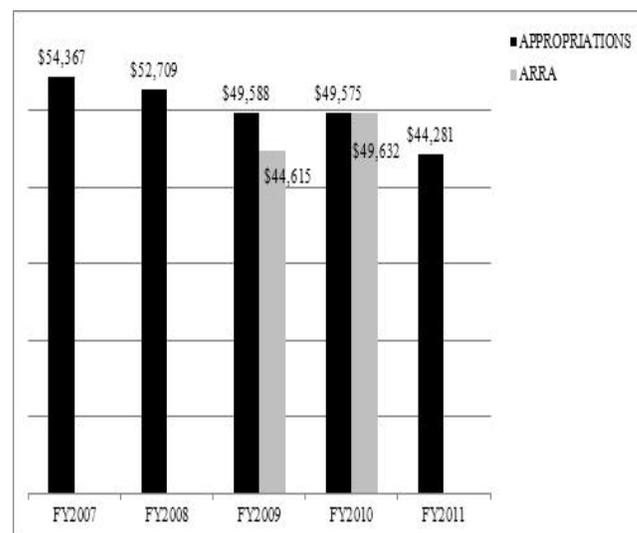
- **Research Support**
- **Resource Support**
- **Career Development Support**
- **Training Support**
- **Small Business Research & Development Support**

## Overview of FY2011

NLM'S EP FY2011 base budget was \$56,463,000 including \$44,281,000 for research funding and \$12,182,000 allocated to contracts for the NN/LM.<sup>[1]</sup> As shown in **Figure 1**, NLM's Appropriation-based grants budget has decreased since 2007. Despite a challenging year, characterized by a substantial budget cut, several continuing resolutions, and a near government shutdown, EP continued to support NLM's mission by funding meritorious biomedical informatics research and resource projects. One hundred forty-four new and continuing grants were awarded to 70 academic organizations located across the country (See **Table 7**). NLM grantees and trainees produced more than 200 peer-reviewed publications, cited 140 times by other researchers. (See **Table 9** for more details).

<sup>[1]</sup> All future reference to EP grant budgets in this report exclude NN/LM funding, which is reported in the Library Operations section of the NLM annual report.

**Figure 1: NLM Grants Budget - FY2007-FY2011** (in thousands)



For the second consecutive year, a new NLM research grantee received a prestigious Presidential Early Career Award for Scientists and Engineers (PECASE). This year's awardee, Dr. John Brownstein of Children's Hospital Boston, received the award for his work on real-time tracking of the global impact of climate change on infectious disease. He is the fifth NLM grantee to receive this honor. Information about NLM's five PECASE winners is available at <http://www.nlm.nih.gov/ep/PECASE.html>.

In 2011, EP staff made major efforts to make information about NLM's grant programs more accessible to the community. Three Webinars were presented, new content was added to the EP Web site, and staff made 18 presentations about NLM's grant programs. In support of NLM's 175<sup>th</sup> anniversary, two special lectures by NLM grantees were presented.

During 2009 and 2010, EP's grants budget swelled by nearly \$90 million of American Recovery and Reinvestment Act (ARRA) funds, including \$17 million from the NIH Office of the Director for NLM-administered ARRA grants. In all, NLM awarded 251 ARRA grants, grant supplements and contracts. With all ARRA funds awarded, the role of EP staff has turned to monitoring and documenting ARRA funded grant activities. A summary of NLM's ARRA funding can be found at <http://www.nlm.nih.gov/recovery/>.

## 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 6** shows success rates in 2009, 2010 and 2011 for NLM's core grant programs for applications funded with appropriated funds. NLM's success rates dropped in 2011 due to budget constraints. Efforts to maintain the Research grant success rate at 15% or better were successful. Other NIH Institutes and Centers also saw drops in their success rates in 2011. [http://report.nih.gov/success\\_rates/index.aspx](http://report.nih.gov/success_rates/index.aspx)

### Research Grants

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

#### Research Grant Program

EP receives R01 research grant applications at three deadlines each year through the NIH parent announcement, NLM's Express grant program, and various multi-institute NIH initiatives in which NLM participates. NLM's research investment areas include basic biomedical informatics, clinical/health informatics; translational bioinformatics, public health informatics and user-centered informatics.

- 83 reviewed R01 applications (85 in FY2010)
- 15 awarded R01 applications (17 in FY2010)

In FY2011, NLM launched a new research grant program titled "NLM Advanced Informatics for Health." This program seeks exceptionally innovative research that can achieve its goals in four years. A total of 18 applications were received in response to the FOA. Review and awards will take place in FY2012.

#### Exploratory/Developmental Grants

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

- 33 reviewed R21 applications (26 in FY2010)
- 3 awarded R21 applications (7 in FY2010)

**Table 6: Success Rates - Core NLM Grant Programs FY2009 through FY2011**

Grants Program	2009	2010	2011
Research Project Grants (R01)	17%	20%	18%
SBIR/STTR Small Business Grants (R41-R44)	8%	11%	13%
Exploratory Research (R21)	9%	27%	9%
Applied Informatics Resource Grants (G08)	10%	15%	7%
Scholarly Works Grants (G13)	14%	20%	9%

### Conference Grants

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

- 3 reviewed R13 applications (4 in FY2010)
- 1 awarded R13 application (1 in FY2010)

### Small Business (SBIR/STTR)

By law, all grant-issuing agencies set aside a portion of available research funds for 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). NLM's SBIR/STTR interests for FY2011 focused on tools for managing interactive publications and/or large datasets, modeling tools for climate and environmental effects on human health, new technologies for disaster information management and tools to enable communities to use health indicators, such as the HHS Health Indicators Warehouse, to improve a community's health. In FY2011, EP received applications through the NIH parent announcement, and through a joint-agency SBIR FOA titled "Robotic Technology Development and Deployment (RTD2)." Three of the funded FY2011 SBIR projects were in the area of robotics.

- 40 reviewed small business applications (24 in FY2009)
- 5 awarded small business applications (5 in FY2010)

## Resource Grants

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

### *Information Resource Grants to Reduce Health Disparities (G08)*

In 2010, NLM suspended its longstanding open announcement for Applied Informatics projects, replacing it with an RFA entitled Information Resource Grants to Reduce Health Disparities (RFA-LM-10-001). The focus of the RFA was projects that bring useful, usable health information to health disparity populations and their health care providers through the use of computer and information science. Applications were reviewed and awards made in 2011. Although the same number of awards was made as in 2010, the success rate fell due to a more than 100% increase in applications.

- 60 reviewed Information Resource Grants to Reduce Health Disparities applications (27 Applied Informatics G08 applications in FY2010)
- 4 awarded Information Resource Grants to Reduce Health Disparities applications (4 Applied Informatics G08 awards in FY2010)

### *Grants for Scholarly Works (G13)*

NLM alone, among the NIH Institutes and Centers, provides grant funds to support the preparation of scholarly manuscripts. The first grants awarded by NLM in 1965 were for history of medicine projects. The Scholarly Works program continues to play a key role in important areas of biomedical scholarship, particularly in the history of science and medicine. In December 2010 NLM re-issued the funding announcement for this program, with the first application deadline set at February 23, 2011. Although the number of awards was restricted as compared to 2010, the success rate fell due to the large increase in applications.

- 57 reviewed Scholarly Works applications (35 in FY2010)

- 5 awarded Scholarly Works applications (7 in FY2010)

## Training and Career Awards

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

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

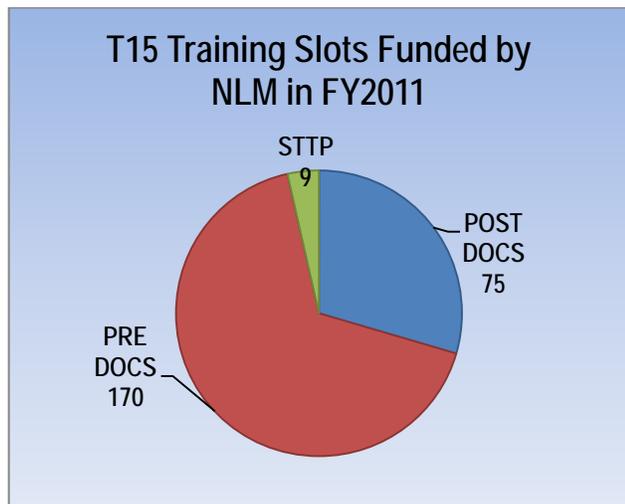
Five-year institutional training grants support pre-doctoral, post-doctoral, and short-term informatics research trainees at 18 university-based programs across the country (see **Figures 2 and 3**). The current five-year period ends in June 2012. RFA 11-001, NLM Institutional Training Grants for Research Training in Biomedical Informatics, was issued in February 2011. Twenty-eight applications were received; they were reviewed in July 2011. Notable in this competitive round was the downsizing of NLM's training program from 18 programs to 15 or less; a ceiling of 15 or less on the number of trainee slots at each program; a shift toward translational bioinformatics that emphasizes a connection to human health.

Collectively, the 18 existing 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 late 2010, 16 training programs received supplemental awards to develop or update curriculum materials. The end products, designed to be shared with other training programs, were discussed at the 2011 NLM Training Conference. Curriculum topics included translational bioinformatics; human phenotyping; global health informatics; clinical information extraction; professional career development; quantitative methods for genomics; and team based learning.

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. Nine STTP slots were awarded for use in FY2011. Programs may appoint STTP trainees at any time during the training year, which runs July 1, 2011 through June 30, 2012.

**Figure 2: FY2011 NLM T15 Training Program (for a total of 254 trainees)**



The annual NLM Informatics Training Conference took place this summer at the National Institutes of Health, Bethesda, Maryland on June 28-30, 2011. Three hundred twenty-three people attended, including directors, faculty, staff, and trainees from all 18 current NLM training programs; faculty and trainees from the Veterans Administration informatics training sites; NLM staff; and invited guests. Trainee research projects were presented in plenary, parallel or open-mic sessions by 55 informatics trainees. An additional 44 trainees presented research-related posters at the meeting. Attendees voted for best speaker, best poster and best open-mic talk. The award winners were:

- Open-Mic: Robert Bruggner of Stanford University, for “Automated Identification of Predictive Cell Populations from High-Dimensional Mass Cytometry Data.”
- Best Poster: Day 1: Daniel Newkirk of UC-Irvine, for “Mechanisms of Cohesion in Transcriptional Regulation of Nipbl (+/-) Mice”; Day 2: Matthew Kayala of UC-Irvine, for “Learning to Predict Chemical Reactions.”
- Best presentation: Day 1: Ryan Rochat of Baylor College of Medicine, “Zernike Phase Contrast Cyro-EM Reveals the Portal Complex in Herpes Virus”; Day 2: Vincent Fusaro of Harvard Medical School, “Designing Effective Clinical Trials using Simulations.”

Because this year marked the 175<sup>th</sup> anniversary of the National Library of Medicine, a special informatics careers panel was held on the final day. Chaired by Ted Shortliffe, President and CEO of the American Medical Informatics Association (AMIA) and former director of

two different NLM training programs, the panel presented six former NLM trainees who discussed their research career path and their work in the field of biomedical informatics. Details of the panel and conference are available at <http://www.nlm.nih.gov/ep/trainingconf2011.html>.

#### *Summer Research Experience Program (R25)*

Building on a similar successful ARRA program, NIH launched a new NIH Summer Research Experience Program, which provides a high quality research experience for high school and college students during the summer academic break. NIH expects that the programs will help attract young students to science careers, and provide opportunities for college students to gain valuable research experience to help prepare for graduate school. NLM participates in this program, but only accepts applications from its University-based informatics training programs (T15). Each awardee expects to train eight people each summer.

- 3 reviewed R25 applications
- 3 awarded R25 applications.

#### **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.

- 12 reviewed K99/R00 applications (3 in FY2010)
- 3 awarded K99/R00 applications (3 in FY2009)

##### *K22 Independent Career Development Award*

In FY2010, NLM re-launched its 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.

- 1 reviewed K22 applications
- 0 awarded K22 applications

##### *Loan Repayment Programs (L30, L40)*

EP ended its participation in the NIH loan repayment programs in FY2011, due to low relevance to our research objectives, and budget constraints.

**Figure 3: In FY2011, 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)

## **Pan-NIH Projects**

### *National Center for Biomedical Computing (NCBC)*

The National Centers for Biomedical Computing are cooperative agreement awards funded under the NIH Common Fund. As a result of an open competition for the second five-year period, seven NCBC centers were funded in 2010. NLM continues to administer one NCBC center, "Informatics Integrating the Bench and Bedside (i2b2)," based at Harvard's Brigham and Women's Hospital, which received four years of additional funding in 2010. Each year, the NIH Common Fund will provide a decreasing share of the funds for i2b2, through 2014. NLM's share of support for i2b2 will increase over the years. By the final year, NLM will provide 100% of the funding for i2b2. NHLBI also provides co-funding for i2b2. NLM program officers have scientific advisory roles in three NCBC centers: i2b2 at Harvard, 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. NLM also selectively participates in topic-focused multi-institute funding announcements. The active multi-institute programs NLM participates in are listed in **Table 8**.

The applications for multi-institute programs are reviewed by CSR. Applications assigned to NLM that receive fundable merit scores are considered for awards alongside grants reviewed by NLM's study section. Links to the multi-institute initiatives in which EP participates are incorporated into the grant programs list on the EP Web site at <http://www.nlm.nih.gov/ep/Grants.html>.

### *Shared Funding for Research & Training within NIH*

In FY2011, NLM provided co-funding support to an OD-NIH (Pioneer Award), an NHGRI (STTR Award), an FIC (International Research Training Award) and the iDASH NCBC. These collaborative funding arrangements represent co-funding by the NLM in the amount of \$761,163. NLM received co-funding from OD-NIH, NHLBI, NIDCR in the amount of \$2,863,834.

### *Interagency Agreements and Funding*

NLM grantee John Brownstein received supplemental grant funding support from the Office for the Assistant Secretary for Preparedness and Response (ASPR), HHS, for a special health-related initiative relating to his HealthMap platform and data sources, aiding development of a tool to provide epidemiologic/disease

outbreak information for hurricane impacted areas, resulting in improved situational awareness and timely analysis of disease threats. [\$100,000]

Along with several other NIH Institutes, NLM provides support to the Protein Sequence Databank (PDB) at Rutgers University. [\$200,000] The PDB is the single world-wide repository for the processing and distribution of 3D biological macromolecular structure data.

### **Extramural Programs Web Site**

<http://www.nlm.nih.gov/ep/index.html>

In 2011, there were 61,340 visitors to the EP Web site, a decrease of 2% over the previous year. The EP Web site continued to have a substantial interest from international visitors who represented almost 19% of total visits. The most frequently viewed pages were those describing the grant programs, EP Home page, information pages such as the NLM 2011 Funding Plan and Grant Deadlines, Frequently Asked Questions, and a page of tutorials for preparing applications.

According to the American Customer Satisfaction Index (ACSI) survey results, the overall satisfaction score for the EP Web site was 70 out of a possible 100, which is similar to the satisfaction ratings for the NLM main site.

### **EP Operating Units**

#### *Program Office*

#### *Grant Program Development*

Program activities in FY2011 were focused on implementation of plans for our regular grant programs, post-ARRA grant funding activities, which ended in FY2010.

#### *Advanced Informatics for Health*

Following successful experiences with the ARRA challenge grants and two rounds of participation in the NIGMS Eureka grant program, NLM decided to launch its own challenge grant program to solicit exceptionally innovative informatics research projects. For the first Advanced Informatics for Health FOA, NLM sought projects to construct an intelligent computer program that offers health advice to a person about a complex health problem, based on knowledge from (1) the person's personal health record (PHR) or electronic medical record (EMR) and (2) knowledge sources in the public domain. Contingent on the availability of grant funds, NLM is planning to issue an Advanced Informatics for Health solicitation every year, with a single receipt date. Each FOA will pose a problem in health or biology for which there is no current widely-adopted informatics solution.

## *Reissued Programs*

EP issued three additional FOAs in FY2011, as follows:

(1) Reissue of NLM's Grants for Scholarly Works in Biomedicine and Health (G13) FOA, which expired in February 2010. This FOA supports the preparation of book-length manuscripts and other scholarly works of value to US health professionals, public health officials, biomedical researchers and health sciences historians.

(2) A request for new and continuation applications for NLM Institutional Training Grants for Research Training in Biomedical Informatics (T15). This RFA invited applications for support of predoctoral and postdoctoral training for research careers in biomedical informatics. A Webinar was held to provide technical assistance and answer questions; nearly 50 prospective applicants participated.

(3) Reissue of NLM Express Research Grants in Biomedical Informatics (R01) FOA, which expired in May 2011. This FOA supports research that advances the science of biomedical informatics. Informatics is concerned with the optimal organization, management, dissemination and use of information.

## *Outreach Activities*

At the NLM Informatics Training Conference, program staff presented a Career Development and New Investigator awards seminar. About 40 people attended. Additionally, the tutorial was offered as a one-hour Webinar titled "NLM Webinar - Introduction to Career Development Awards and ESI/NI R01 Grants in Biomedical Informatics," on Sept. 22, 2011. The goal of the session was to provide information to prospective applicants about our career development awards and research project grants. Nearly 100 attendees logged in for the Webinar. A transcript of the presentation and a list of Q&A's have been posted at the EP Web site.

EP organized a series of Informatics Lectures by NLM grantees as part of NLM's 175<sup>th</sup> anniversary events. Dr. Dean Sittig presented on "Clinical Decision Support: What is it? Why is it so hard? What can we do about it," on March 2, 2011, at Natcher on the NIH campus. Dr. Eneida Mendonca presented on "Bringing Scientific Evidence into Clinical Practice: Challenges, Successes and Failures," on June 8, 2011, at the same venue. These lectures were well attended in person and via videocast, and were archived for later viewing.

## *Grants Management Office*

The NLM issued 104 noncompeting grant awards and 40 competing grant awards in FY2011. The total base grants budget was \$44,281,000. The EP/NLM grants office provides the daily and end-of-fiscal year accounting for the EP budget including all awarding mechanisms. NLM

grants staff continues to maintain the NIH – Financial Conflict of Interest Database and respond to Freedom of Information Act requests for NLM extramural grants.

At the NLM Informatics Training Conference, grants management staff and DEAS staff assigned to EP presented an in-person session for training grant administrators, answering questions and providing a tutorial on the xTRAIN system. At the request of the attendees, a listserv was established to facilitate continued dialogue on administrative topics that affect all programs.

## *Scientific Review Office*

### *Grant Review Activities*

Overall, 357 applications were reviewed for which NLM was the primary assignment. Of those, 298 were reviewed by NLM. The remaining 59 applications were reviewed by CSR including SBIR/STTR grant applications.

### *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 FY2011 and reviewed 118 applications (as compared to 116 in 2010). The Committee reviews applications for biomedical informatics and bioinformatics research applications received through the NIH parent or NLM-issued funding announcements, career transition awards, and some special issuances such as the Advanced Informatics for Health.

### *Special Emphasis Panels (SEPs)*

Eight Special Emphasis Panels were held during FY2011 compared to six in FY2010. 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, or when a direct conflict of interest exists between the application and a member of the BLIRC. A special panel was assembled in July 2011 to review T15 institutional training grants (the program is competed only once every five years). The panel, composed of 20 well-regarded experts in the biomedical informatics training field, reviewed 28 applications. Overall, NLM's SEPs reviewed a total of 180 applications during FY2011, compared to 115 in FY2010.

### *Review Concurrence*

Concurrence with the results of initial review, called second-level review, is performed by the Board of Regents (BOR). The BOR Extramural Programs Subcommittee conducts early concurrence reviews electronically on the most fundable research grants and on special initiatives such as the Training Grant RFA. In

FY2011, the subcommittee held three early concurrence panels; 57 grants were voted on and approved. The BOR conducts an en bloc vote for all other applications assigned to NLM as primary or secondary Institute. For

the fiscal year, a total of 1880 NLM grant applications were included in the en bloc votes (312 primary and 1568 dual).

**Table 7: Extramural Programs Grants Budget by NIH Mechanism Groupings, and by Activity Code-FY2011**

<i>(dollars in thousands)</i>		
FY 2011 operating budget request by NIH mechanism groupings	FY 2011 Budget	
	Total	
	No.	Amount
Research Project Grants (R01, R03, R21, R00, RL1, U01)	75	\$ 24,495,234
SBIR/STTR (R41, R42, R43, R44)	6	\$ 821,552
Other Research - Research Careers (K99)	6	\$ 632,517
Other Research - Biomedical Research Support (P41)	4	\$ 2,024,616
Other Research - Other (G08, G13, R13, R25, D43)	33	\$ 2,694,993
Training - Institutional (T15)	18	\$ 12,283,942
R&D Contracts (L30, L40, N01, Y03)	1	\$ 200,000
National Centers for Biomedical Computing Award (U54)	1	\$ 1,128,146
EP budget excluding TAPS and RMS/Direct Operations	144	\$ 44,281,000
FY 2011 operating budget request by activity code	FY 2011 Budget	
	Total	
	No.	Amount
D43: International Training Grants in Epidemiology (cofund)	-	\$ 235,163
DP1: Pioneer Award	-	\$ 396,000
G08: Knowledge Management & Applied Informatics; Planning Grant for	11	\$ 1,335,646
G13: Scholarly Works in Biomedicine and Health	16	\$ 874,057
K99: Pathway to Independence	6	\$ 632,517
L30: Extramural Loan Repayment Program	-	\$ -
L40: Extramural Loan Repayment Program	-	\$ -
N01: NN/LM Contracts	-	\$ -
P41: Biomedical Resource Grant	4	\$ 2,024,616
R00: Pathway to Independence	9	\$ 2,049,758
R01: Research Project Grants	54	\$ 20,247,324
R03: Small Project Grants	1	\$ 66,150
R13: Conference Grants	3	\$ 65,084
R15: Academic Research Enhancement Award (AREA)	-	\$ -
R21: Exploratory/Developmental Grants	10	\$ 1,736,002
R25: Education Projects	3	\$ 185,043
R41: Small Business Technology Transfer (STTR)	-	\$ -
R42: Small Business Technology Transfer (STTR)	-	\$ 80,000
R43: Small Business Innovation Research (SBIR)	6	\$ 741,552
R44: Small Business Innovation Research (SBIR)	-	\$ -
RL1: Linked Research Project Grant	1	\$ -
T15: University Biomedical Informatics Research Training Programs	18	\$ 12,283,942
U01: Cooperative Agreement	-	\$ -
U54: NCBC Roadmap Center	1	\$ 1,128,146
Y03: Inter-Agency Agreement	1	\$ 200,000
EP budget excluding TAP and RMS/Direct Operations	144	\$ 44,281,000

Figure 4: FY2011 EP Grants budget (dollars in thousands)

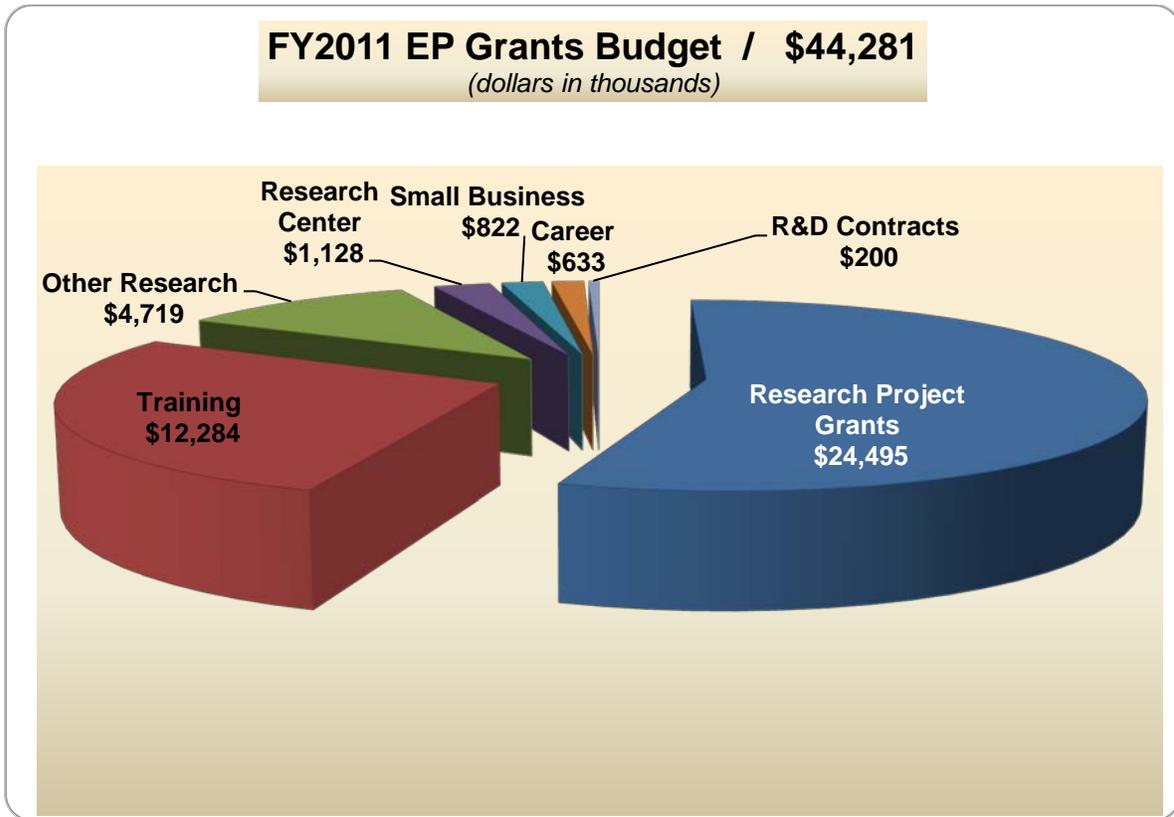


Figure 5: FY2011 Research Grant Investment Areas, \$6.75 million

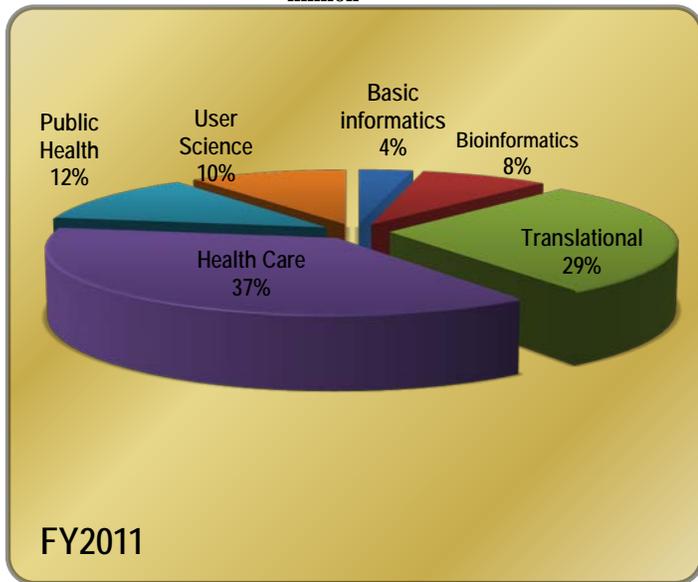
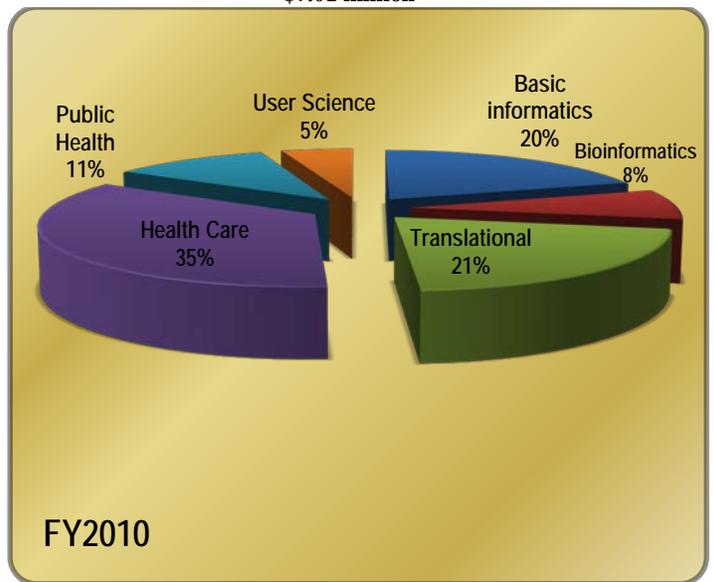


Figure 6: FY2010 Research Grant Investment Areas, \$7.01 million



**Table 8: RFA/PA Actions in FY2011**

<b>NLM's Core Active Grant Programs</b>		
<b>Announcement</b>	<b>Title</b>	<b>Expiration</b>
PAR-11-208	NLM Express Research Grants in Biomedical Informatics and Bioinformatics (R01)	8-May-14
PAR-11-260	Research Project Grants (NIH Parent R01)	8-May-14
RFA-LM-11-002	NLM Advanced Informatics for Health (R01)	26-Jul-11
PA-11-261	Exploratory/Developmental Grants for Biomedical Informatics and Bioinformatics (NIH Parent R21)	8-Jan-14
PA-10-071	NLM Informatics Conference Grants (R13)	8-Jan-13
PA-10-070	Academic Research Enhancement Award (AREA) Research Grants (R15)	8-Jan-13
RFA-LM-10-001	NLM Information Resource Grants to Reduce Health Disparities (G08)	Expired on July 2010
PAR-11-084	NLM Grants for Scholarly Works in Biomedicine and Health (G13)	24-Feb-12
PAR-10-195	NLM Independent Career Development Award for Biomedical Informatics (K22)	8-May-13
PA 11-197	NIH Pathway to Independence Award (K99/R00)	8-Jan-14
<b>Multi-IC Active Announcements in which NLM Participates</b>		
<b>Announcement</b>	<b>Title</b>	<b>Expiration</b>
PA-08-190	Research Supplements to Promote Diversity in Health-Related Research	30-Sep-12
PA-08-191	Supplements to Promote Reentry into Biomedical and Behavioral Research Careers	30-Sep-12
PAR-10-133/135	Understanding and Promoting Health Literacy Research Grants (R01)(R21)	8-May-13
PAR-10-136/137	Behavioral and Social Science Research on Understanding and Reducing Health Disparities Research Grants (R01)(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
RFA-AG-11-010 RFA-HL-11-033/034/035	NIH Basic Behavioral and Social Sciences Opportunity Network (OppNET) Grants	10/14/2010 01/06/2011
PA-11-096	Small Business Innovation Research (SBIR) Grant (R43/R44)	8-Jan-12
PA-11-097	Small Business Technology Transfer Grants (STTR) (R41/R42)	8-Jan-12

### Publications in FY2011

In FY2011, NLM-funded projects led to the publication of more than 200 articles, which received 140 citations during FY2011. Ninety four of the articles were written by NLM trainees. Table 4 shows the journals in which NLM grantees and trainees publish most frequently.

**Table 9: Top Journals in which NLM Grantees and Trainees Published in FY2011**

<i>Journal</i>	<i>Grantees</i>	<i>Trainees</i>
Journal of the American Medical Informatics Association (JAMIA)	22	7
Bioinformatics	12	0
PloS one	9	4
Journal of biomedical informatics	8	5
BMC bioinformatics	8	5
Clinical pharmacology and therapeutics	5	2
Medical image computing and computer-assisted intervention (MICCAI)	5	0
BMC genomics	4	2
Nucleic acids research	4	0
Journal of dental education	3	0

**Table 10: Five Most Frequently Cited Articles Acknowledging NLM Funding in FY2011 (Scopus)**

Publications	Citations
Apweiler, R., Binns, D., Stanley, E., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Staehli, S. <b>Ongoing and future developments at the Universal Protein Resource</b> (2011) <i>Nucleic Acids Research</i> , 39 (1), pp. D214-D219.	34
Wilke, R.A., Savova, G., Xu, H., Denny, J.C., Roden, D.M., Krauss, R.M., McCarty, C.A., Lamba, J. <b>The emerging role of electronic medical records in pharmacogenomics</b> (2011) <i>Clinical Pharmacology and Therapeutics</i> , 89 (3), pp. 379-386.	10
Dudley, J.T., Butte, A.J. <b>In silico research in the era of cloud computing</b> (2010) <i>Nature Biotechnology</i> , 28 (11), pp. 1181-1185.	9
Hunninghake, G.M., Barnes, K.C., Wilk, J.B., Oconnor, G.T., James Gauderman, W., Vora, H., Baurley, J.W., Mathias, R.A. <b>TSLP polymorphisms are associated with asthma in a sex-specific fashion</b> (2010) <i>Allergy: European Journal of Allergy and Clinical Immunology</i> , 65 (12), pp. 1566-1575.	8
Pearson, J.F., Bachireddy, C., Shyamprasad, S., Goldfine, A.B., Brownstein, J.S. <b>Association between fine particulate matter and diabetes prevalence in the US</b> (2010) <i>Diabetes Care</i> , 33 (10), pp. 2196-2201.	7

**Table 11: NLM New Grants Awarded in FY2011**

(Sorted by PI name, within each grant category)

**RESEARCH GRANTS**

BURNSIDE, ELIZABETH S  
1 R01 LM010921-01A1  
UNIVERSITY OF WISCONSIN MADISON  
Integrating Machine Learning and Physician Expertise for Breast Cancer Diagnosis

CHAPMAN, WENDY W  
1 R01 LM010964-01  
UNIVERSITY OF CALIFORNIA SAN DIEGO  
Interactive Search and Review of Clinical Records with Multi-layered Semantic Annotation

DENNY, JOSHUA CHARLES  
1 R01 LM010685-01A1  
VANDERBILT UNIVERSITY  
From GWAS to PheWAS: Scanning the EMR phenome for gene-disease associations

GOPALAKRISHNAN, VANATHI  
1 R01 LM010950-01  
UNIVERSITY OF PITTSBURGH AT PITTSBURGH  
Bayesian Rule Learning Methods for Disease Prediction and Biomarker Discovery

HURDLE, JOHN F  
1 R01 LM010981-01A1  
UNIVERSITY OF UTAH  
POET-2: High-performance computing for advanced clinical narrative preprocessing

LI, XIAOBAI  
1 R01 LM010942-01  
UNIVERSITY OF MASSACHUSETTS LOWELL  
New Technology to Preserve Patient Privacy and Data Quality in Health Research

LIU, GILBERT C  
1 R01 LM010923-01  
INDIANA UNIV-PURDUE UNIV AT INDIANAPOLIS  
Delivering Geospatial Intelligence to Health Care Professionals

LIU, YIN  
1 R01 LM010022-01A2  
UNIVERSITY OF TEXAS HLTH SCI CTR HOUSTON  
Bayesian Methods in Signal Transduction Network Analysis

LU, XINGHUA  
1 R01 LM011155-01  
UNIVERSITY OF PITTSBURGH AT PITTSBURGH  
Ontology-Driven Methods for Knowledge Acquisition and Knowledge Discovery

MOONEY, SEAN DAVID  
2 R01 LM009722-06A1  
BUCK INSTITUTE FOR RESEARCH ON AGING  
Informatic profiling of clinically relevant mutation

PAGE, C. DAVID  
1 R01 LM011028-01  
UNIVERSITY OF WISCONSIN MADISON  
Secure Sharing of Clinical History & Genetic Data: Empowering Predictive Pers. Medicine

PESTIAN, JOHN P  
1 R01 LM011124-01  
CHILDREN'S HOSPITAL MEDICAL CENTER CINCI  
Multi-Institutional Pediatric Epilepsy Decision Support

SALISBURY, KENNETH  
1 R01 LM010673-01A1  
STANFORD UNIVERSITY  
Patient-Specific Simulation for Surgical Rehearsal

SALTZ, JOEL H  
1 R01 LM011119-01  
EMORY UNIVERSITY  
Informatics for Integrative Brain Tumor Whole Slide Analysis

YE, JIEPING  
1 R01 LM010730-01A1  
ARIZONA STATE UNIVERSITY-TEMPE CAMPUS  
Computational Methods for Expression Image Analysis

**KNOWLEDGE MANAGEMENT/APPLIED INFORMATICS GRANTS**

CALMAN, NEIL S  
1 G08 LM011056-01  
INSTITUTE FOR URBAN FAMILY HEALTH  
Implementing MedlinePlus Connect in Spanish to Address Health Disparities in Spanish-Speaking Communities

CHUNG, BOWEN (New Investigator)  
1 G08 LM011058-01  
LA BIOMED RES INST/ HARBOR UCLA MED CTR  
DREW\_UCLA  
Connect Community Partnered Resources to Improve Depression Outcomes

PIRIE, PHYLLIS L  
1 G08 LM011075-01  
OHIO STATE UNIVERSITY  
Reaching a wider audience: Modifying Net Wellness  
content and outreach

TOOEY, MARY JOAN (New Investigator)  
1 G08 LM011079-01  
UNIVERSITY OF MARYLAND BALTIMORE  
NLM Information Resource Grants to Reduce Health  
Disparities

### SCHOLARLY WORKS

BLANC, PAUL D  
1 G13 LM010076-01A1  
UNIVERSITY OF CALIFORNIA SAN FRANCISCO  
New Technologies, Novel Diseases; Industrial Illness in  
20<sup>th</sup> Century Rayon Manufacturing

GALLAGHER, CATHERINE A (New Investigator)  
1 G13 LM010936-01  
LLOYD SOCIETY, INC.  
Health Policies and Outcomes for Criminal Justice: A  
resource for policy makers, practitioners and researchers.

SUTTER, PAUL SHRIVER (New Investigator)  
1 G13 LM010906-01  
UNIVERSITY OF COLORADO AT BOULDER  
Pulling the Teeth of the Tropics: An Environmental and  
Public Health History of the Construction of the Panama  
Canal

WEBB, JAMES LEWIS ADRIAN  
1 G13 LM010888-01  
COLBY COLLEGE  
Malaria in Africa: A History of Infections and  
Interventions, 1900-2010

WOLF, JACQUELINE HELENE (New Investigator)  
1 G13 LM010878-01  
OHIO UNIVERSITY ATHENS  
A Social History of Cesarean Section in the United States

### CAREER DEVELOPMENT

BEKHUIS, TANJA CATHARINA (New Investigator)  
1 K99 LM010943-01A1  
UNIVERSITY OF PITTSBURGH AT PITTSBURGH  
Screening Nonrandomized Studies for Inclusion in  
Systematic Reviews of Evidence

CAMI, AUREL B (New Investigator)  
1 K99 LM011014-01  
CHILDREN'S HOSPITAL BOSTON  
Mapping the Drugome: Predictive Network Approaches to  
Drug Safety Surveillance

FUSARO, VINCENT A  
1 K99 LM011020-01  
HARVARD UNIVERSITY (MEDICAL SCHOOL)  
Repurposing Electronic Health Data for Serendipitous  
Therapy Discovery

### CONFERENCE GRANTS

WONG, BANG T.  
1 R13 LM010987-01  
BROAD INSTITUTE, INC.  
VizBi: A Conference on Visualization in Biology

### NIH PATHWAY TO INDEPENDENCE AWARD

BEKHUIS, TANJA CATHARINA (New Investigator)  
1 K99 LM010943-01A1  
UNIVERSITY OF PITTSBURGH AT PITTSBURGH  
Screening Nonrandomized Studies for Inclusion in  
Systematic Reviews of Evidence

CAMI, AUREL B (New Investigator)  
1 K99 LM011014-01  
CHILDREN'S HOSPITAL BOSTON  
Mapping the Drugome: Predictive Network Approaches to  
Drug Safety Surveillance

FUSARO, VINCENT A  
1 K99 LM011020-01  
HARVARD UNIVERSITY (MEDICAL SCHOOL)  
Repurposing Electronic Health Data for Serendipitous  
Therapy Discovery

### SMALL PROJECT GRANTS

LEROY, GONDY (New Investigator)  
1 R03 LM010902-01  
CLAREMONT GRADUATE UNIVERSITY  
Large-Scale Evaluation of Text Features Affecting  
Perceived and Actual Text Difficulty for Online Health  
Information Consumers

### SMALL BUSINESS INNOVATION RESEARCH (SBIR) AND SMALL BUSINESS TECHNOLOGY TRANSFER (STTR) AWARDS

BAJPAL, GAURAV (New Investigator)  
1 R43 LM011328-01  
TECHNO-SCIENCES, INC.  
Development of a clinical robotic device for diagnosis,  
rehabilitation and treatment-prediction of knee injuries

BUNIN, BARRY A (New Investigator)  
1 R43 LM011152-01  
COLLABORATIVE DRUG DISCOVERY, INC.  
Biocomputation across distributed private datasets to  
enhance drug discovery

JUDKINS, TIMOTHY N (New Investigator)  
1 R43 LM011326-01  
INTELLIGENT AUTOMATION, INC.  
Development of a mobile robot with an affective interface  
and human activity tracking toward socially assistive  
interaction with the elderly

MERCHANT, ZAFFER (New Investigator)  
1 R43 LM011325-01  
AVENTUSOFT, LLC  
Speech Therapy Robot (STR) to assist in the  
administration of evidence based speech and language  
therapy

ROSIN, CHRISTOPHER D (New Investigator)  
1 R43 LM011099-01  
EBM PARITY COMPUTING, INC.  
Automated matching of relevant research studies to patient  
records for EBM

## **NIH SUMMER RESEARCH EXPERIENCE PROGRAMS**

BALDI, PIERRE  
1 R25 LM011170-01  
UNIVERSITY OF CALIFORNIA IRVINE  
UCI BIT Undergraduate Summer Research (BIT-SR)  
Program

GADD, CYNTHIA S (New Investigator)  
1 R25 LM011174-01  
VANDERBILT UNIVERSITY  
Vanderbilt Biomedical Informatics Summer Research  
Experience Program

HERSH, WILLIAM R  
1 R25 LM011168-01  
OREGON HEALTH AND SCIENCE UNIVERSITY  
OHSU Summer Internship in Biomedical  
Informatics for College Undergraduates

# OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

*Ivor D'Souza*  
*Director*

The Office of Computer and Communications Systems (OCCS) provides efficient, cost-effective computing and networking services, application development, technical advice and collaboration in informational sciences. OCCS provides some NLM services directly, and also indirectly supports NLM research and management programs.

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

Brief discussions of this year's activities in the following subject matter areas will be presented:

- Consumer and Public Health
- Controlled Medical Vocabularies
- IT Infrastructure Services
- Medical Literature Support and Document Delivery Services
- Outreach and Customer Services

## Consumer and Public Health

### *MedlinePlus Mobile and MedlinePlus*

MedlinePlus Mobile's monthly drug update process was modified to include the display for the "unapproved" drug market status. The XML site maps for MedlinePlus Mobile (English and Spanish) were released to Google, Yahoo and Bing, which will enhance the Search Engine Optimization (SEO) performance.

There were two releases of MedlinePlus Mobile in FY2011, in February and in June. These releases included:

- Changes to the URL structure to closely match the URL of MedlinePlus' full site.
- Enhancements to the search function to make it more secure.
- Addition of preferred links to all mobile pages.
- An updated drug page view to display the American Society of Health System Pharmacist (ASHP) new 'market-status' data.

- Changes to permit more effective access by Web crawlers (Google, Yahoo, etc.).

Three versions of MedlinePlus were released this year, in March, June and July. These releases included:

- That all pages are XHTML compliant.
- Share buttons and RSS buttons were added on news-related pages.
- Enabled easier public access to the MedlinePlus Health Topic RSS feeds by linking them from the "RSS Feeds" static page, and from all Health Topic pages, and enabling RSS browser auto-detect.
- Enhanced the RSS Feeds static page ensuring that it will be up-to-date with the most current topic group categories, topic groups, and health topic names.
- Added text transcripts to the HealthDay TV video public page.

Due to decreasing usage and the availability of similar information through other online resources, NLM and its partners have phased out the MedlinePlus Go Local project. The last three Go Local areas of Alabama, Indiana and Michigan have been removed, as well as the input system. All Go Local codes and references have been removed from the MedlinePlus system.

MedlinePlus saw over 700 million page views and over 172 million unique visitors. Additionally, MedlinePlus and Spanish MedlinePlus received a satisfaction score of 87 and 86 respectively out of a possible 100 on the ACSI E-Government Satisfaction Index, a survey that is one of the most comprehensive and representative reflections of the citizen experience with federal government Web sites. They ranked third and fourth in the Information/News category among federal Web sites. Over 124 production requests have been implemented this year.

### *MedlinePlus Connect*

MedlinePlus Connect is a service which provides a way for public entities to request the URL of MedlinePlus topic pages that are a match to the ICD9/SNOMED CT codes they supply. It now also serves up MedlinePlus Web pages related to medications in the same manner by using RxNorm Concept Unique Identifiers (RxCUI) and National Drug Codes (NDC). MedlinePlus Connect was selected as one of three "Secretary's Picks" for the HHS *Innovates* Challenge on March 31. During mid-April, the service expanded to include querying Lab Tests via the LOINC Code System. A new feature was added to include the patient handout links to the Web services.

### *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 over 7,500 Structured Product Labels (SPL).
- Created and deployed a mobile Web site for DailyMed.
- Added new search features to the DailyMed Web site. In this new search version, customers can retrieve labels based on Pharmacological class – such as Mechanism of Actions, Physiologic Effects, Therapeutic, and Pharmacokinetics.
- Added SPL Imprint Data and SPL Manufacturer data to RxNorm attributes.
- Published more than 2,500 unapproved drug labels.

Use of DailyMed increased by 67 percent from 70 million page views to over 117 million page views by 5 million unique visitors.

#### *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. Fifty topics are now available in NIHSeniorHealth, two new topics were added this year, Long Term Care, and Anxiety Disorders. Audio review and text-to-speech pronunciation adjustments were completed for several topics. The Accent module received numerous enhancements including updated speech dictionary with tunings from voice font vendor Loquendo.

NIHSeniorHealth received over 13 million site hits and over 538,000 unique visitors, over 25 percent of whom were international visitors.

#### *NLM Digital Repository Project*

The NLM Digital Repository Project supports collection and preservation of a wide variety of digital objects, including manuscripts, pamphlets, monographs, images, movies, audio and other items. The repository includes digitized representations of physical items, as well as born digital objects. OCCS provides system architecture and software development resources to assist in the implementation and maintenance of the NLM digital repository. During FY2011, a key new repository functionality was implemented. All films in the repository have Section 508 captions and transcripts, and are fully searchable with the unique OCCS-developed Video Search software.

On September 26, HHS Secretary Kathleen Sebelius named the NLM Video Search software one of

six winners in Round 3 of the HHS *Innovates* competition; the innovation was included in the honorable mention awardees. 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 DC Fedora Users Group, and to representatives from the National Agricultural Library.

To provide high repository availability, a full Active-Active redundancy configuration for Digital Collections was implemented at the Bethesda and offsite data centers. Each location has two repository production servers and a full set of the repository data. Also, four new software releases were promoted to the production servers. These releases added new user interface features, improvements to the book and video data models, and improvements for constructing collections. OCCS also developed a prototype for periodically testing the data integrity of master preservation files, and implemented a Google XML Sitemap for the repository, which will help Web search engines index all the repository content.

For the Medical Heritage Library (MHL) Project, OCCS defined workflows for pre-ingest and ingest processing for the 6,000 books that NLM will digitize for the MHL project, and developed software scripts that automate many of the workflow steps. OCCS created a software “pipeline” that has enabled hundreds of MHL books to be processed each month; nearly 1,200 books were processed in September. During FY2011, OCCS created the new “Medicine in the Americas” collection in Digital Collections for MHL books, and ingested 1,378 books into the production servers. Nearly 1,800 additional books are in the pipeline for ingest in FY2012.

OCCS also developed a Web-based quality assurance system for the NLM Digitization Program to automate, improve and streamline the quality assurance processes. The first books were processed through the new interface on January 27. Since then 3,427 books have been processed with over 2 million files of pages and metadata.

There were two new collections, 18 new videos, over 1,400 new books, and approximately 300,000 page images added to the production repository. Roughly 1.8 million files were used in ingesting new books and videos. In addition, Digital Collections recognized more than 30,000 unique visitors and over 395,000 page views to its site.

#### *Health Services Research Projects in Progress (HSRProj)*

The HSRProj Web site was updated with over 2,000 new projects, which increased the total number of projects available for searching to 22,994. Additionally, 1,204 records with the Final Date between 1/2006 and 12/2006 were relocated to the archive file. The HSRProj Search was redesigned in order to fit the new NLM Header. Download action was improved by creating pre-built pages for “complete” and “tagged” format. A beta of Vivisimo search was added as an option.

### *PHPartners and HSR Info Central*

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

The modifications made to PHPartners.org and the HSR Info Central Input System included successfully promoting version 9 into production, which included a new 'Archived' table where links from expired 'Web Pages' table will be relocated. Also, necessary changes were made to accommodate HP 2020 (Healthy People 2020) Initiative data. The changes implemented to HSR Info Central included adding information to the PHPartners site regarding Healthy People 2020 Initiative.

### **Controlled Medical Vocabularies**

#### *Unified Medical Language System (UMLS) Project*

The Unified Medical Language System Metathesaurus is a large multipurpose, multilingual vocabulary database that contains information about biomedical and health-related concepts. The 2011AA edition of the UMLS Metathesaurus includes 155 source vocabularies in 21 languages. The Metathesaurus contains over 2.4 million concepts, a 9 percent increase from last year (with nearly 10.7 million names for those concepts, an 8 percent increase), along with over 50 million relationships between and among these concepts. This represents a nearly nine percent growth in relations since the 2010AA edition of a year ago. In the past two release cycles (2010AB and 2011AA), more than 30 English language and 18 translated sources were successfully updated or added. The SPECIALIST tools and the lexicon were also updated in the 2011AA edition.

A newly designed common technology platform for supporting UMLS applications is now in use, which acts as the underlying object model for MetamorphoSys, UMLS Terminology Services (UTS), 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

released, replacing the legacy Knowledge Source Server (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 browser for the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT). OCCS personnel reconfigured the download site to improve UTS authenticated file distribution. The UTS system is operating 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 FY2011, significant systems infrastructure was brought online to support IHTSDO-related efforts within NLM, which included subversion servers, a TeamForge installation, a Maven repository and a build server all set to authenticate against UTS user accounts. This infrastructure has enabled us to manage internal workflow efficiently to support outside collaborators in a variety of ways.

With international collaborators, OCCS made significant progress toward the development and deployment of tooling used to support all aspects of workflow management for the IHTSDO project to map SNOMED CT to ICD-10—from creation of initial batches through the quality assurance and publication of mapping data in release format version 2 (RF2).

Using the technology, methodology and some of the data from the ICD-10 mapping project, progress was made in a new effort to map SNOMED CT to ICD-10-CM for use within the US. Two editions of the US Extension to SNOMED CT were successfully produced, validated and published.

Four Kaiser Convergent Medical Terminology (CMT) subsets were received, validated and published, which included CMT Top 2500, CMT Cardiology, CMT Mental Health, and CMT Musculoskeletal.

The US SNOMED CT Content Request System was released for public use. This content request system supports a variety of different request types, form-based interactive creation of requests as well as batch uploading via a structured Excel spreadsheet template, and manager functions to move requests through a predefined workflow, eventually modeling them in the next US Extension edition. Training Webcasts were developed and recorded to assist users in interacting productively with the system.

## *RxNorm Project*

RxNorm, produced by NLM, is a standardized nomenclature for clinical drugs. Seven major versions of the RxNorm Editing System were released this fiscal year, in October, January, February, March, June and September. Implementation of these releases included RxTerms implementation support, disassociate brand name module, concept report module, attributes of attributes and dose form groups module, and merge process tracking. They are a benefit to users of the data as well as the users of the editing system. For the data users, there is now “one-stop” shopping for data from RxNorm and RxTerms. The editors have numerous new features that assist them in creating and linking RxNorm data, and have a way to see what merges have been performed in the system.

There were 12 monthly Full releases, four monthly Current Prescribable Content releases, 52 Weekly releases, and nine Current Prescribable Content Weekly Update releases of RxNorm this year. OCCS completed the monthly inversion and insertions of nine regular sources, a total of 264 source insertions, new inversions of MMX and NDDF datasets for branded drugs, NDC codes and other attributes. There were 254 requests for system changes that were processed. Two resynchronizations were completed with UMLS data. Among other changes, 73 new quality assurance checks were implemented within the release process.

RxNorm currently contains 20,356 active generic drugs (a 5 percent increase from FY2010), 17,386 active branded drugs (an 8 percent increase from FY2010), 6,566 active ingredients (a two percent increase from FY2010), 15,388 active brand names (a 10 percent increase from FY2010), and 250,238 distinct NDC codes for RxNorm forms (a 15 percent decrease from FY2010). The decrease is due to removal of NDC codes from drugs no longer on the market. This is an important tool for e-prescribing.

## *Medical Subject Headings (MeSH) and Related Systems*

MeSH is the NLM’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:

- Cutover for the French, German, Italian, Czech and Croatian translations. Prepared the test database for the Korean translation.
- Cutover for the 2012 MeSH. Generated and distributed the new 2012 MeSH XML for various translations.
- Loaded the new OMIM (Online Mendelian Inheritance in Man) terminology into the internal MeSH Database.
- Implemented the new Global Citation Management System (GCMS) functionality and Interface for

processing Publication Types and Other DCMS Elements (ODE) Manual Tasks.

## **IT Infrastructure Services**

### *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. Recurring cycles of vulnerability scanning, detection, and remediation continue to ensure concrete improvements in the NLM’s security posture.

OCCS coordinated an NLM-wide security assessment and penetration test to evaluate the effectiveness of security controls implemented on NLM IT systems, in collaboration with the NIH Office of the Chief Information Office (OCIO). The assessment team conducted a comprehensive network scan looking at network security and exposure to attacks. NLM performed exceptionally well in this exercise, and was commended for its high-caliber IT security program.

Several security tools were replaced this year to improve our awareness of and responsiveness to new security vulnerabilities.

- NLM contracted for a distributed denial-of-service (DDoS) mitigation protection through an outside vendor. A DDoS attack occurs when multiple systems flood the bandwidth or resources of a targeted system. The NLM/Internet traffic is constantly monitored for data containing potential DDoS signatures, and if detected, NLM can switch its data traffic to a filtering network for removal of the harmful DDoS traffic. This system was tested via a live, staged test, and was verified to work under controlled conditions.
- To participate in accomplishing NIH and HHS information security management goals, NLM collaborated with NIH on the implementation of the first major HHS enterprise-wide security solution, HHS’ Distributed Security Event and Information Management (SEIM) System for real-time detection and alerting and long-term security log analytics.
- NLM adopted an enterprise application scanning tool, which drives security best practices into existing software development lifecycles. It will assist applications developers in identifying and remediating application vulnerabilities earlier in the development process.
- NLM adopted a firewall configuration management tool to automate and enforce consistency of firewall changes.
- A privilege management tool was adopted to automate temporary security privilege elevation to NLM personnel who need elevated privileges to conduct specific tasks.

Due to the increase of new vulnerabilities and the rapid emergence of associated threats at the desktop, more

software patches than ever before must be deployed, and with a much greater degree of urgency. More than 1,000 desktop computers were updated with over 14 million signatures, a 65 percent increase from FY2010. These patches were deployed overnight to avoid user interruption and minimize downtime.

Anti-virus screening continued for all outgoing Web (HTTP) connections to the Internet, blocking over 1,500 security violations, and minimizing the burden on desktop security software to fend off these threats.

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 FY2011 Security Awareness Training for employees, contractors and fellows.

### *High Speed Communication Networking*

OCCS continued to provide reliable and secure LAN and Internet communications services to support the various programs across NLM. Steps were taken to increase the capabilities and reliability of network services and storage by providing for the following:

- Data communications services for the NLM offsite computing facility in Sterling, Va.
- 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 from 1 Gigabit-per-second (Gbps) to 2 Gbps, with a 1 Gbps backup circuit. These circuits connect 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 Sprint and Time Warner. This failover capability is tested once per month.

A new Internet service contract was awarded to Global Crossing in September 2011. The new service, anticipated for activation in January 2012, will be for 3 Gbps, and can be quickly upgraded to 10 Gbps if needed.

Internet2 is an important resource for connection with NLM and a consortium of universities and research communities. Internet2 connectivity is a 10 Gigabit link to the high-speed backbone network via the Mid Atlantic Exchange (MAX) at the University of Maryland. A redundant, diverse 10 Gbps fiber connection from NLM to the MAX is provided by FiberGate. Access to the

Commercial Peering Service (CPS) was added through MAX in 2011, which allows NLM to transfer data with peering partners at very high data speeds.

The NLM perimeter network provides a 10 Gigabit security boundary to aggregate connections to NIHnet, Internet, Internet2 and the offsite Data Center. The "Interconnect Layer" provides for 10 Gigabit connections to OCCS, LHC and NCBI divisional networks. A private, dark fiber connection has been deployed from the NLM main campus site to NLM satellite sites - EP (in Rockledge) and SIS, OCCS and OAM (in Democracy 1 and 2). This network bandwidth was upgraded from 1 Gbps to 3 Gbps. The offsite Data Center's Internet2 connection bandwidth was upgraded from 622 Mbps to 2.5 Gbps (1 Gbps currently usable).

State-of-the-art firewalls guard the network at the perimeter and lower layers within the local NLM network. The NLM perimeter and NCCS firewalls run on 10 Gigabit appliances, providing high-capacity throughput to support NLM's high volumes of Internet and Internet2 traffic.

NLM was one of the few entities in the Federal government that demonstrated its readiness to deliver Internet services over the next-generation IP protocol, IPv6. NLM successfully participated in the World IPv6 Day (June 8, 2011) by making the NLM Home Page, MedlinePlus, and DailyMed Web sites available via the IPv6 protocol.

A new remote access/Secure Socket Layer (SSL) Virtual Private Network platform was procured and implemented in 2011. This replaced the Cisco VPN concentrators that were end-of-life, and provided new capabilities that simplify remote access connectivity for NLM remote staff and teleworkers.

Network availability was provided at more than 99.98 percent and network services availability at 99.99 percent or higher.

### *Data Center Reengineering*

Working with the NLM Office of Administration, OCCS took many significant steps to advance the safety, reliability and performance requirements of the NLM Data Center. These included:

- Expanded Data Center space by an additional 2,500 square feet.
- Increased electrical power capacity by 50 percent from 900 kW to 1.35 MW.
- Installed an electrical busway distribution system that has reduced the time taken to add new electrical circuits from a few months to a few hours.
- Installed an environmental monitoring software system that detects water leakage in the Data Center and automatically alerts Data Center personnel to take corrective action.
- Increased the infrastructure of In-Row Coolers (IRC) from 40 to 77. The IRCs will create a high density zone that will allow power usage in excess of 12 kW/s

per cabinet – allowing NLM to pack twice as many systems in a high-density computer rack on the Data Center floor.

- Installed six 30-ton air conditioners to maintain a constant 70 degree temperature. This additional capacity allows NLM to sustain the failure of two air-conditioner units and still maintain operations.

### *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 from 306 to 460 virtual machines, a 50 percent increase from FY2010. Physical server counts trended down in FY2011 with a 31 percent reduction, from 265 down to 184. The increase in virtual servers and decrease in physical servers maximizes the use of available computing resources.
- Removed 57,000 linear feet of electrical cables from the NLM Data Center sub floor thereby increasing the cubic foot per minute (CFM) air movement by 25 percent allowing computer cabinets to operate at a higher wattage capability.

Desktop virtualization continued with virtual machines (VM) being integrated into desktop operations. Use of a VM gives users flexibility without requiring them to maintain a second physical computer. In addition, several software applications were virtualized, such as Remedy, Oxygen, and Delpro. 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.

### *Business Continuity and Disaster Recovery*

In order to protect NLM's mission-critical systems the Center for Information Technology (CIT) and NLM implemented an off-site Data Center, called the NIH Consolidated Colocation Site (NCCS), in Sterling, Va., in 2004. This off-site Data Center serves as a disaster recovery/alternate computing site for NLM, other NIH institutes and centers, and other HHS organizations.

This year, the off-site Data Center's Internet2 bandwidth was upgraded from 622 Mbps to 1 Gbps. 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 the off-site Data Center 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. Virtual

Private Networking and Citrix remote access are implemented at the off-site Data Center. Various other upgrades were performed to the storage systems and servers located at this site.

### *Office Automation and Desktop Support*

OCCS introduced support for two Lifecycle Work Stations (LWS) from which secure changes to NIH smart cards, such as PIN resets and certificate renewals, can be performed more conveniently. NLM staff and contractors located in the Democracy Suites and surrounding areas are able to complete required smart card updates without traveling to the main campus.

NLM migrated to Internet Explorer 8 (IE8) after internal testing and compatibility testing with over 90 applications.

Since the 2003 Help Desk consolidation with NIH's IT Help Desk, NLM desktop and PC networking support requests are 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 74 percent of the calls (9,119 tickets) with 6 percent of support calls being completed by NIH staff.

Incident management procedures were updated to improve communication with technical points of contact and the NLM user community during major incidents affecting access to various networked services.

### *Web Analytics*

NLM uses the WebTrends software package to track the number of pages served over time by the sites being managed and to provide detailed analysis of trends in site usage, audience composition, and other matters. This year a new data collection method called page tagging was deployed on a selected number of NLM Web sites (Intranet, NLM main Web home page, NLM databases). This method is a Web analytics industry standard that gives richer and more accurate usage data. NLM will be using this method to improve insights on how online visitors navigate the NLM Web sites. For instance, WebTrends link tracking was used on the NLM home page to measure changes made during the NLM main Web redesign to highlight further areas for improvement.

### *Web Support*

Rate throttling was implemented to support NLM's Emergency Access Initiative (EAI). Rate throttling helps prevent wholesale automated download of the resources provided by EAI. OCCS' work on preventing unauthorized access to EAI publications has reassured the publishers participating in EAI.

## *Search Engine*

Support continued for the Vivisimo production search implementations. In addition to upgrading Vivisimo to version 8, the following initiatives were undertaken:

- A nightly automatic refresh of drug synonyms was implemented, improving the search quality in our NLM main Web search and MedlinePlus searches.
- The Digital Repository search was completed and released, which indexes XML files and other resources on the NLM Digital Collections Web site, and provides search functionality.
- The NLM FindIt search for internal staff was completed and released which facilitates data retrieval against the NLM Intranet Web site, a shared network drive, several NLM SharePoint sites, and all NLM wiki spaces. This application implements social search features to allow users to rate, tag and comment on the search results. These features make it possible for users to share their search experience and help us to improve the search quality.

## *NIH Combined Federal Campaign (CFC) Web site*

OCCS was responsible for creating and maintaining the NIH CFC Web site for this year's campaign. The Web site was created and deployed on September 30. Also, a photo contest application was created for viewing and voting in support of the NIH IC Directors' Challenge.

## **Medical Literature Support and Document Delivery Services**

### *Data Creation and Maintenance System (DCMS)*

The major event for the Indexing Data Creation and Maintenance System (DCMS) this year was 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 19.5 million records, an eight percent increase in the amount of data exported over last year. Over 49,000 "new" OLDMEDLINE records were loaded and processed for the publication year 1946. More than 1.5 million instances of subject terms on OLDMEDLINE records were mapped to current MeSH terms to improve their accessibility in Medline/PubMed. Also:

- The Potential Abstract Labels (PAL) process finds and saves potential abstract label strings. OCCS developed rules to find the label strings with the most potential and a new Web based tool to record their decisions. From over 52,000 unique potential abstract label strings, this helps narrow the number to be considered to less than 2,000, expediting label selection and entry into DCMS.
- DCMS now fully supports Publisher Data Review System (PDRS). Lister Hill Center's PDRS currently has three tools to assist DCMS data review staff in

performing their tasks. The PDRS "Investigators tool" allows the user to cut and paste long lists of investigators from the online article, and the tool parses and passes the data to DCMS. The other two tools search PubMed Central articles for Databanks & Grants, and passes this data back to DCMS.

- A new version of the DCMS Loader & Extractor was completed and placed in production which included enhancements to the data dictionary, cursor performance, logging, parallel queries, new jobs, process control locking and testing.
- As part of the annual baseline extraction of Meeting Abstracts, Lister Hill Center's (LHC) Medical Text Indexer (MTI) provided more than 121,000 records loaded and re-released to the Gateway system, a 6 percent increase over the prior year.
- Over 4 million new "Cites" (new reference type) were loaded and exported from DCMS. There are currently over 36 million Cites.
- Over 21,000 duplicate citations were received from the PubMed Central project, which were identified and deleted.

## *DOCLINE*

DOCLINE, the NLM interlibrary loan (ILL) system, supported approximately 2,498 domestic and international libraries in processing more than 1.7 million interlibrary loan transactions and 300,000 Loansome Doc requests. A new major version of DOCLINE was released in July, and several other minor releases were also promoted. The Resubmit and Loansome Doc Transfer modules were redesigned for more efficient work flow and to match the architecture and interface of the other modules. These were the final modules in the interface redesign project. New features include a "free" link to items in the NCBI BookShelf, the ability to select an ISBN if multiple ISBN's have been assigned to an NLM Unique ID, and maintaining holdings for NLM's National Print Retention Program. Seven libraries have committed to retain 25 titles in the program. There were 70 DOCLINE and Loansome Doc enhancement requests implemented in addition to internal security and server configuration changes. DOCLINE users ordered 899,738 distinct PubMed articles, articles from 22,410 different journals, and 13,973 monographs.

## *Voyager Integrated Library System (ILS)*

The year-end processing of the Voyager Integrated Library System database was completed in November 2010. Approximately 13,000 records were modified to convert MeSH headings to the 2011 versions. There were changes to other fields to improve the integrity of our data, which included the imprint, series statement, special producer and other fields. Baseline files of all records were created in both MARC and XML formats. These included more than 3.6 million records from four different files. NLM installed

release 8 of Voyager on a Linux platform; the previous versions of Voyager have been on Solaris machines.

#### *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. The main achievement for this year was the processing of the eTK/eVK data for loading into ENCompass. Also, programs were written and tested to load the data.

#### *Print Journal Donation Tracking System*

An automatic system to track the print journal donations was implemented. This system facilitates donations of material missing in the NLM collection from libraries, etc. using DOCLINE holdings if available. Four hundred thirteen journal donations have been submitted. There are 1,603 total titles offered, 4,387 total volumes offered and 1,209 titles received.

#### *Literature Selection Technical Review Committee (LSTRC)*

An application enhancement was implemented to support administrators in assigning users and reviewers to access the application, and for reviewers to have the ability to add and update their review forms online. Modifications made to display completed review forms to show information once all the review forms have been completed. Implemented programs to display a report that includes all the titles listed for a particular review date along with the UID, URL and username/password for each journal title and the data from Serials and Medline Review application. The report can be saved in an Excel file. In addition, created Completed Review Forms view after all titles are completed by reviewers.

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 update data from publishers in some tables if the data in Serials/LSTRC are updated, a link was added to view only applications to the publisher on the home page, and added a validation message on submit so that the publishers/editors only submit the same application in one day.

#### *Images from the History of Medicine (IHM)*

The Images from the History of Medicine system provides access to the nearly 70,000 images in the prints and photograph collection of the History of Medicine Division (HMD) of the US NLM (NLM). The collection includes portraits, pictures of institutions, caricatures, genre scenes and graphic art in a variety of media, illustrating the social and historical aspects of medicine.

- Updated software and promoted the new Luna version 6.3 system to the public.
- Conducted vulnerability scanning of the test IHM site.
- Loaded new images into Luna and set up processing scripts on the Luna production environment.

#### *HMD Directory*

The HMD Directory is an online version of the History of Medicine component of DIRLINE (Directory of Information Resources Online), a NLM (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 included adding an icon to save the collections in PDF file format and modifying some programs in order to add the NLM header and footer on the public site. The Emergency Preparedness and Response mobile Web site was converted into an Apple iOS application for HMD.

#### *Serials Extract File (SEF)*

The Serials Extract File is a critical internal product that serves as a bridge between other applications. Among numerous upgrades and fixes, a new ColdFusion Manage Serials Form was promoted into production, which replaced an older version that used Oracle Forms technology. The new form includes better application security, easier site navigation; new features aimed at increasing user efficiency and a friendlier user interface than the previous version. Two forms, Manage Users and Manage Problem Authors were developed using ColdFusion software for DCMS. Although the Manage Users form replaced an older Oracle Forms version, the Manage Problem Authors form is a new application. It allows DCMS users to correct reported mistakes in author names. The 2011 List of Serials Indexed (LSI) publication was generated in XML format for the NLM ftp server.

#### **Outreach and Customer Services**

OCCS staff continued to support the NLM outreach program through volunteering and providing technical support to over 30 outreach events. In addition, OCCS provided active support for several NLM exhibition initiatives:

### *Native Voices Exhibition*

OCCS provided Web site maintenance, support and edits as well as provided first line support (software) for Interview Station Kiosks in the Native Voices exhibition. OCCS is primarily responsible for adding new content, edits and troubleshooting and for providing software troubleshooting support for iPad installation at the “Coffee Bar” as needed.

A Vivisimo search application for Native Voices was implemented which facilitates data retrieval from the Native Voice Exhibition Web site as well as retrieves timeline and interview videos based on their metadata. This application provides the ability to limit the search to a certain category (era/theme/region). A spelling suggestion feature was also implemented. Three pieces of software were integrated with Vivisimo search:

- Onsite Interactive Timeline Kiosks use Vivisimo to search a collection of events on a timeline.
- Onsite Interactive Interview Kiosks use Vivisimo to search through a collection of interviews and interview transcripts.
- The exhibit’s static html Web site uses Vivisimo searches in three different contexts requiring sophisticated faceting and filtering.

### *Web Exhibit*

Web exhibit activities include building a new design interface with Outreach Activity Reporting Form (OARF) in order to support Web exhibit and consumer health. The Requirements Refinement Process and the interface design are being worked on in parallel.

### *Against the Odds Web Exhibition and Web Interactive Activities*

The main accomplishment this year was the conversion of all externally hosted content for the exhibit and converting it into HTML files. New files were moved into the locally hosted site, and the links were rewritten as necessary. The Rotating features on the home page was reworked so that these features work in a random manner, instead of linked to a particular month.

### *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. The application was upgraded to ColdFusion 9/Oracle 11g. Loan form and condition reports were created. Functional improvements included the addition of reports, the enhancement of search features that allow users to search and conduct operations on the search set without having to re-query, and the option to use predefined filters to better support the product workflow.

### *Customer Service Support System (Siebel)*

Multiple defect resolutions, enhancement releases and system upgrades were delivered for the Customer Service, Change Request and Firewall Service Request Management applications. Key accomplishments included:

- Completed the Siebel 8.1.1.1a release enhancement and moved it into production.
- Tested and moved the new ILL (Interlibrary Loan) Web form request into production.
- Supported customer service to process 36,016 service requests, 63,613 inbound e-mail messages, 50,741 outbound e-mail messages, 596 firewall change requests, 600 application change requests, and 73 source acquisition requests.

### *Bethesda Hospital Partnership*

The Bethesda Hospitals’ Emergency Preparedness Partnership (BHEPP) was formed between the National Naval Medical Center, National Institutes of Health Clinical Center, Suburban Hospital Healthcare System, and the NLM. BHEPP’s main goals include:

- Rapid and successful response to any emergency situation.
- Integration of the collaborative BHEPP response with other community, regional and national responses.
- Creation of collaborative structures and processes to serve as an exportable model for other similar centers in the Nation.

OCCS led BHEPP research in the areas of data and voice communication infrastructure to facilitate cooperation and collaboration among BHEPP member sites, enhancing security features by encrypting patient data and related documents, and adding triage category and transfer patient counters on the Real Time Location System (RTLIS) dashboard for Incident Command Center (ICC) use during a mass casualty incident.

# ADMINISTRATION

Todd D. Danielson  
Associate Director for Administrative Management

**Table 12: Financial Resources and Allocations, FY2011**  
(Dollars in Thousands)

<u>Budget Allocation</u>	
<b>Program Area</b>	<b>Amount</b>
Extramural Programs	\$62,503
Intramural Programs	259,975
Library Operations	(78,068)
Computer and Communications Systems	(29,531)
Lister Hill National Center for Biomedical Communications	(47,227)
National Center for Biotechnology Information	(86,796)
Specialized Information Services	(18,353)
Research Management and Support	14,255
Total Appropriation	336,733
Plus: Reimbursements	54,781
<b>Total Resources</b>	<b>\$391,514</b>

## Personnel

In October 2010, **Leonardo Mariño-Ramírez, PhD**, was appointed to a staff scientist position at the National Center for Biotechnology Information (NCBI) after having been in Colombia for two years. Dr. Mariño-Ramírez earned a doctorate in biochemistry and biophysics from Texas A&M University for work in the development of a genetic system to study protein-protein interactions, funded by a Fulbright predoctoral fellowship. He carried out postdoctoral training in Dr. David Landsman's laboratory at NCBI. Dr. Mariño-Ramírez focuses on the development of an efficient framework to study various aspects of gene regulation in eukaryotic genomes using integrative computational tools and epigenetic data. He is also an editor of *GENE* (Elsevier) and *DATABASE* (Oxford Journals). As a staff scientist, he will continue this work along with other projects.

In October 2010, **Alex Gavino, MD**, arrived as a postdoctoral fellow at Lister Hill National Center for Biomedical Communications (LHNCBC). Dr. Gavino has his medical degree from University of the Philippines, Manila where he was the Coordinator for Telehealth Services of the UP Manila National Telehealth Center. He will be working with Paul Fontelo on mobile telehealth projects involving processing and sending of telehealth referrals through mobile phones and/or digital camera.

In October 2010, **Robert Jenders, MD**, joined the Lister Hill Center as a staff scientist. Dr. Jenders received his MD degree from the University of Wisconsin-Madison,

where he also completed his internship and residency in medicine. He went on to serve as clinical and research fellow in medicine at the Massachusetts General Hospital, completing a National Library of Medicine fellowship in medical informatics at the Harvard-MIT Division of Health Science and Technology. Prior to joining the Lister Hill Center, Dr. Jenders was clinical professor of medicine and attending physician in the Department of Medicine at Cedars-Sinai Medical Center (CSMC) at UCLA. His research and operational work focus on clinical decision support systems, electronic health records and medical vocabularies. Dr. Jenders has served as co-chair of the clinical decision support work group of HL7, the premier standards development organization for health information technology (HIT), for over a decade. He will work with Dr. McDonald on the PHR project and data standardization efforts including vocabulary and message standards.

In November 2010, **Jennifer Marill** was appointed chief of the Technical Services Division (TSD) in Library Operations. Jennifer has been the deputy chief of TSD since March 2007, and has worked in NLM's Public Services Division from 1998-2003. From 2003-2007, she worked at the Office of Strategic Initiatives at the Library of Congress (LC). At both NLM and LC, Jennifer has been successful in defining and managing projects and in applying good organizational principles to work performance. Jennifer received her BA in international studies from Johns Hopkins University, her MA in Russian and east european studies from the University of Michigan, and her MLIS from the University of Illinois. Jennifer will manage the staff responsible for selecting, acquiring, processing, and cataloging print and digital items for our collection. TSD staff also coordinates all journal licensing activities, implements policies related to collection development, and has responsibility for the repository of digital monographs and other formats that complement PubMed Central.

In November 2010, **Raheleh Salari, PhD**, started her research fellow position at the National Center for Biotechnology Information (NCBI) after she earned a doctorate in computer science from the Simon Fraser University for work in computational biology. Before joining the lab for computational biology at SFU, she graduated with an MSc in computer science from Sharif University of Technology, Iran in 2005. In the summer of 2009, she visited the group of Rolf Backofen at the University of Freiburg, working on sparsification of RNA-RNA interaction. During her PhD, she mainly focused on ncRNA gene finding and RNA structure and interaction prediction. Her research interests comprise RNA structure and interaction, protein-protein interaction network, protein sequence alignment and protein homology search. As a research fellow, she will work on Protein-DNA binding discovery along with other projects on disease network.

In November 2010, **George R. Riley, PhD**, was appointed to a staff scientist position at the National Center for Biotechnology Information (NCBI) after working at the AABB. Dr. Riley earned a doctorate in biology from Georgetown University for work on developmental regulation of glycoprotein synthesis. He has done research on developmental genetics at the University of Washington and on the molecular biology of RNA editing at Seattle Biomedical Research Institute. He has been a forensic DNA laboratory director, has developed DNA laboratory quality assurance and quality systems, and has worked at the AABB as a laboratory quality systems assessor. As a staff scientist he will work to build a computerized quality assurance tool for assessing and interpreting measurements of short tandem repeat (STR) markers, initially for forensic DNA typing, as well as other projects.

In November 2010, **Lu Sun, PhD**, was appointed to a staff scientist position at the National Center for Biotechnology Information (NCBI) after having been a ComputerCraft contractor since 2007 at NCBI. Dr. Sun earned a doctorate in entomology from Purdue University for research on insect systematics and phylogenetics. He then worked for Purdue University as a post doctoral associate. He has engaged in research and education of identification, taxonomy and evolution of organisms, primarily insects and nematodes, with approaches involving character research, gene sequence analysis, computing, and also paleobiology and biogeography. He joined ComputerCraft in December 2010, as a contractor for NCBI, to build and maintain the NCBI taxonomy database with a focus on the invertebrate section and provide consults towards nomenclature, up-to-date classification and other taxonomy-related issues generated in submissions. As a staff scientist, he will continue this work.

In December 2010, **Stefan Jaeger, PhD**, arrived as a visiting scientist at Lister Hill National Center for Biomedical Communications (LHNCBC). Dr. Jaeger received his doctorate in computer science from University of Freiburg, Germany in 1998. Dr. Jaeger has an international research background, both in industry and academics. He has held positions at Daimler Research Center Ulm, University of Karlsruhe, Tokyo University of Agri. & Tech., University of Maryland, Chinese Academy of Sciences and Max Planck Society. Prior to coming to NLM, he was a visiting scholar at University of Missouri-Columbia. His research interests include pattern recognition, machine learning, artificial intelligence, and information fusion. He will be working with Sameer Antani in biomedical imaging.

In January 2011, **Janice E. Kelly** was appointed chief, Outreach and Special Populations Branch in the Division of Specialized Information Services. For the past 17 years she was executive director of the National Network of Libraries of Medicine, Southeastern/Atlantic (SE/A) Region at the University of Maryland, Baltimore. Prior to

that, she served as associate director of the NN/LM Greater Midwest Region. She has directed many outreach initiatives involving libraries and community agencies, expanded the educational programs of the SE/A and managed an award program targeting health professionals and consumers in support of the mission of the NN/LM. Previously, Ms. Kelly held positions in academic health sciences libraries, a hospital library, and multi-type library network. She received her BA in secondary education from the University of Pittsburgh and her MLS also from the University of Pittsburgh. Ms. Kelly manages the staff responsible for outreach programs, a number of specialized Web sites, promotional activities, and the SIS exhibit program.

In February 2011, **Ye Wang, MS**, was appointed to a staff scientist position at Lister Hill National Center for Biomedical Communications (LHNCBC) after obtained a MS degree (2008) in computer science from the Old Dominion University. He also holds an MS degree (2000) in Computer Science and a BS degree (1997) in computational mathematics from Nanjing University, China. He has been a key contributor on the NLM Personal Health Record (PHR) project in the Computer Science Branch of LHC since 2006, serving as a senior software engineer, and has been responsible for developing many of its best features, including the spreadsheet functionality.

In February 2011, **Matthew S. Simpson, PhD**, began work as a postdoctoral fellow at Lister Hill National Center for Biomedical Communications (LHNCBC) after completing his doctorate in computer engineering at the University of Maryland, College Park. He had been a student trainee at NLM since October 2007. His research interests include biomedical informatics, information retrieval, and the reliability and security of computer systems. He will work with Dr. Dina Demner-Fushman, developing image indexing methods for enabling large-scale biomedical information retrieval using both text and visual features.

In May 2011, **Michael F. Huerta, PhD**, was appointed NLM Associate Director for Health Information Programs Development. Dr. Huerta earned his doctorate in Anatomy from the University of Wisconsin-Madison. He is an internationally known scientist and scholar who has played a key role in the development of the fields of neuroinformatics and imaging. Prior to joining the NLM, Dr. Huerta served as associate director of the National Institute of Mental Health (NIMH) for scientific technology research; director, NIH National Database for Autism Research; director, Office of Cross-Cutting Science and Scientific Technology, NIMH; and associate director, Division of Neuroscience and Basic Behavioral Science, NIMH. Dr. Huerta led the NIMH extramural and intramural efforts related to research and development of technologies, including scientific informatics. Throughout his tenure at NIH, Dr. Huerta's expertise has been

specifically sought by the Office of the Director, NIH for a variety of highly visible and important initiatives such as the Human Brain Project, the National Database for Autism Research, the NIH Blue Print for Neuroscience Research, Small Business Research, and the Human Connectome Project, among others. Dr. Huerta's familiarity with many diverse research activities across NIH, informatics-related and otherwise, will provide this position with the catalyst necessary to expand program efforts and increase collaborations with other ICs and external stakeholders. Dr. Huerta has received numerous awards from both inside and outside the NIH in recognition of his creativity, stellar accomplishments and outstanding leadership. He has authored or co-authored more than 40 publications including scientific papers published in numerous peer-reviewed journals, chapters in scholarly books, and scientific reviews.

In May 2011, **Olivier Bodenreider, MD, PhD**, was appointed as the chief of the Cognitive Science Branch (CgSB), Lister Hill National Center for Biomedical Communications (LHNCBC). Dr. Bodenreider holds an MD from the University of Strasbourg and a PhD in Medical Informatics from the Henri Poincaré University. He also has an MS in Computer Science and an MS in medical information for Hospitals. Dr. Bodenreider has been a staff scientist at LHNCBC for more than 10 years, and has developed, studied and/or contributed to many NLM strategic projects including the Unified Medical Language System (UMLS), the Indexing Initiative, and the Medical Ontology Research (MOR) project. More recently, he has implemented RxNav, a popular system for browsing and using RxNorm and related resources. He is a prolific writer who publishes regularly in the medical informatics literature, and he has substantial experience with clinical vocabularies and natural language processing. Dr. Bodenreider is a superb example of a physician-scientist who has applied his medical and computer science training to important and practical research goals. His creativity, vision, and problem-solving talent will be a major asset as he leads the CgSB's research and development programs. A major goal of his tenure will be to further strengthen the Branch's natural language processing capabilities in order to facilitate many of the Library's strategic goals.

In July 2011, **Joyce Backus** was selected as the Deputy Associate Director for Library Operations. An NLM employee since 1985, Joyce has held a series of progressively challenging positions in the Public Services Division, including systems librarian, section head, and since 2007, deputy chief. In addition to her administrative experience, Ms. Backus has been directly involved in NLM Web development initiatives, the implementation of MedlinePlus and MedlinePlus Connect, and the selection and development of automated systems. She has been active in the Medical Library Association and has represented NLM on various groups dealing with Web

services, electronic health records, and comparative effectiveness. Ms. Backus received her BA from Duke University and her Master in Library Science degree from the Catholic University of America.

In July 2011, **Beth Weston** was appointed as the deputy chief of the Technical Services Division (TSD) in Library Operations. She assists in managing the staff responsible for selecting, acquiring, licensing, and cataloging print and electronic materials for the NLM collection. TSD staff also implement and manage systems supporting these activities. Ms. Weston has been head of the Serial Records Section in TSD since 2002. In this capacity she has successfully planned and administered a coordinated program of acquisitions and licensing for over 20,000 print and electronic serials with annual expenditures exceeding \$7.5 million. From 2000-2002 she was the acquisitions unit head in the Serial Records Section. Prior to her time at NLM, Ms. Weston held positions at two large research libraries, the George Washington University Gelman Library from 1996-2000, and the University of Delaware Library from 1993-1996. Beth received her BA in linguistics from the University of California at San Diego, and her MLIS from the University of California at Berkeley.

In July 2011, **Ivor D'Souza** was selected as Director, Information Systems and Director, Office of Computer and Communications Services (OCCS). He has served in the position of chief, Systems Technology Branch, OCCS, since joining NLM in 2004, and has managed our IT infrastructure operations. Over the past year, Mr. D'Souza also served as acting director, OCCS. He has performed exceptionally well in these important roles. His outstanding leadership, analytical ability, and management skill have helped to keep the library on the leading edge of computing and communications technology. He is highly regarded within the NIH CIO community and has been a key player supporting NLM's participation in the Bethesda Hospitals Emergency Preparedness Partnership and related research projects. Prior to joining NLM, Mr. D'Souza was vice president and partner for AAC, Inc., where he managed IT consulting services provided to major customers including many public, private, commercial and non-profit organizations in the Washington, DC area. Ivor holds a masters degree in electrical engineering from Catholic University of America and a bachelor of science degree in electronics and communications engineering from Mangalore University in India.

In June 2011, **Hamid Haidarian-Shahri, PhD**, joined the Lister Hill National Center for Biomedical Communications (LHNCBC) as a postdoctoral fellow. Dr Haidarian-Shahri received his doctoral degree in computer science from the University of Maryland, College Park, MD in 2011, where his research work focused on data interoperability and semantic search on the Web. His research interests are in the use of database and knowledge

representation techniques to manage and organize information on the Web. Dr. Haidarian-Shahri is working in the Communications Engineering Branch (CEB) on semantic search and visualization strategies for large document collections involving medical informatics. He is mentored by Dharitri Misra, PhD, of CEB.

In June 2011, **Soumi Ray, PhD**, joined the Lister Hill National Center for Biomedical Communications (LHNCBC) as a postdoctoral fellow. Dr. Ray received her doctoral degree from the Department of Computer Science and Electrical Engineering at University of Maryland, Baltimore County in May 2011. For her dissertation, she developed machine learning algorithms to discover and characterize hidden features in time series data. Her research interests are in the fields of Machine Learning, Artificial Intelligence, Time Series, and Reinforcement Learning. She joined the Communications Engineering Branch (CEB) and is working on Automated Metadata Extraction and Web dissemination of scanned medical document collections. She is mentored by Dharitri Misra, PhD, of CEB.

In July 2011, **Michael J. Cairelli, MD**, joined the Lister Hill National Center for Biomedical Communications (LHNCBC) as a postdoctoral fellow and is working in the Cognitive Science Branch (CGSB). Dr. Cairelli graduated from Philadelphia College of Osteopathic Medicine – Georgia Campus in 2011. He also has masters degrees in chemistry and computer science. His research interests are in natural language processing and systems theory, and their application for medical discovery. Dr. Cairelli is working on the application of graph theory to predications extracted by SemRep from MEDLINE citations to identify potential biomarkers for mild traumatic brain injury. He is mentored by Thomas Rindfleisch, PhD, of CgSB.

#### *NLM Associate Fellows Program for 2011-2012*

The NLM Associate Fellowship Program is an annual post-master's program for recent graduates of library/information science 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. Four new Associate Fellows began their year at NLM in September 2011.

**Bethany Harris** received her Master of Science in Information (MSI) degree with a specialization in library and information science from the University of Michigan in 2011. While completing her degree, she worked as a university library associate at Taubman Health Sciences Library, where she provided instruction and reference services. She was a recipient of the Midwest Chapter of Medical Library Association (MLA) annual meeting

scholarship and the Michigan Health Sciences Libraries Association student scholarship in 2010. Prior to graduate school, Ms. Harris spent three years as the manager of a behavioral genetics lab and had previously worked as a pharmacy technician at Lawrence Memorial Hospital. Ms. Harris received her Bachelor of Science (BS) degree in ecology and evolutionary biology from the University of Kansas. While at KU, she participated in two National Science Foundation (NSF) Research Experiences for Undergraduates programs and worked as a research assistant in a quantitative genetics lab.

**Michele Mason-Coles** received her Master of Library Science (MLS) degree from St. John's University, Queens, New York, in 2011, where she attended as a recipient of the Institute of Museum and Librarian Services' (IMLS) Laura K. Bush 21<sup>st</sup> Century Librarian Program scholarship. In 2010, Ms. Mason-Coles was awarded the MLA Minority Scholarship Meeting Award, which provided her the opportunity to attend the annual MLA conference. While completing her degree, she interned at the Weill Cornell Medical Center's Myra Mahon Patient Resource Center. Recently, Ms. Mason-Coles served as a cancer information specialist at the National Cancer Institute's Cancer Information Service, based at Memorial Sloan-Kettering Cancer Center in New York. She received her BS in community health education from Hunter College in New York City.

**Suzy Roy** received her MLS degree from Indiana University, Indianapolis, in 2011 and a Master of Science degree in experimental psychology from Morehead State University in 2005. She has over seven years of neuroscience research experience working in various laboratories, most recently at Purdue University. Her research background spans fields from behavioral neuroscience to molecular and cellular biology. During her MLS training, Ms. Roy was an intern at the Medical Libraries and the School of Nursing at Franciscan St. Elizabeth Health in Lafayette, Indiana, where she worked on creating library research instructional tutorials. She also cataloged reference services, research instruction in School of Nursing courses, and DOCLINE requests. Ms. Roy received her undergraduate degree at Keene State College, Keene, New Hampshire, where she majored in experimental psychology.

**Jessi Van Der Volgen** received her Master of Library and Information Science (MLIS) degree in 2011 from the University of Wisconsin-Madison, where she was awarded the John R. Koch Scholarship. While there, she worked as a reference assistant at Ebling Library for the Health Sciences and at the College Library. Ms. Van Der Volgen was a project assistant at the Limnology Library and completed a Library and Information Literacy Instruction practicum at UW-Madison. Prior to obtaining her master's degree, she was a high school biology teacher in Beaverton, Oregon and was awarded a Partners in

Science grant for genetics research at the Oregon National Primate Research Center. Ms. Van Der Volgen received her BS in life science from the University of Portland.

#### *Retirements and Deaths*

In April 2011, **Becky J. Lyon** retired from her position as Deputy Associate Director of Library Operations, at the NLM, ending 38 years of federal government service. She made significant contributions to NLM, particularly in the Library's outreach to health professionals, public libraries, and patients and their families; emergency preparedness and disaster planning; space planning; and digitization. Ms. Lyon started her NLM career in 1972 as an NLM Associate Fellow, and then held positions in the Technical Services Division and the Lister Hill National Center for Biomedical Communications. She moved into management, first at the Library of Congress, where she headed network services for the National Library Service for the Blind and Physically Disabled, then at the Department of Veterans Affairs, where she directed the Veterans Affairs Library Network (VALNET). Ms. Lyon returned to NLM in 1984 and spent 15 years leading the National Network of Libraries of Medicine. During that time, she played a major role in the expansion of NLM's outreach initiatives to health professionals and the general public. Ms. Lyon became deputy associate director of Library Operations (LO) in 1999. She has been NLM's representative to the Association of Research Libraries, the American Association of Academic Health Libraries, the International Federation of Library Associations, and the European Association for Health Information and Libraries. Ms. Lyon has received numerous honors, including the NLM Director's Award and the National Institutes of Health Merit Award.

In August 2011, **Julia Royall** retired from federal service after 14 years at NLM. She was recruited to NLM in 1997 to create a malaria research communications network (MIMCom) to support scientists in Africa as part of the Multilateral Initiative on Malaria. In 2001, Ms. Royall was made chief of NLM's Office of International Programs. Focusing on Africa, she created programs which engaged and empowered medical librarians, medical journal editors, researchers, clinicians, medical students, healthcare workers and people at the village level for better health. Ms. Royall spent 10 months, in 2007-8, in Uganda as a Fulbright Scholar. She was based at Makerere University, Kampala, Uganda, where she worked closely with medical students, professors, researchers, and librarians. She will remain deeply involved in supporting medical education and research in Africa. A first step will be to work with the MEPI (Medical Education Partnership Initiative), focusing on the role that students and district and village health professionals can play in improving health infrastructure, and on providing ways in which the universities can share their successes and failures and truly create a South-South dialog.

In September 2011, **Dr. Martin M. Cummings** passed away. Dr. Cummings was the NLM Director from 1964 to 1983. Under Dr. Cummings' leadership, NLM's mission and programs were greatly expanded to include the Lister Hill Center intramural research program, Specialized Information Services, the Extramural Grant Program, the National Network of Libraries of Medicine, and the international MEDLARS Centers, and the Library launched MEDLINE and many other online databases. A tribute to Dr. Cummings is located on NLM's Web site at [http://www.nlm.nih.gov/news/cummings\\_death\\_notice.html](http://www.nlm.nih.gov/news/cummings_death_notice.html). An NLM Web exhibition, *Martin Cummings, MD, and the National Library of Medicine: Documents Online*, opened in early November 2011.

#### *Awards*

The NLM Board of Regents Award for Scholarship or Technical Achievement is awarded to recognize and stimulate independent creativity leading to scholarly and/or technical achievements that enrich biomedicine. The recipient of the 2011 award was **Dr. Daniel Le** of the Lister Hill Center Communications Engineering Branch, in recognition for developing the publisher data review system to automatically extract bibliographic data from full-text articles, frequently missing in citations sent to NLM by publishers.

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 2011 award was **Dr. Michael Feolo** of NCBI, for expert project management skills in establishing dbGaP as the most comprehensive single source of information from NIH-funded genome-wide association studies, and **Marina Rappoport** of the Index Section, Library Operations, in recognition of advancing the use of automation to assist in the MEDLINE indexing process through significant contributions to the 2010 MTI Experiment.

The NLM Director's Honor Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to **Hua Florence Chang** in recognition of substantial contributions for the successful coordination, supervision, and development of the many SIS IT products and services including collaborations with ASPR, NCI, and CDC to develop REMM, and **Jane Rosov**, Bibliographic Services Division, LO, for substantial efforts that have led to improvements of the licensing programs at NLM, and for leadership that effects wide-spread use of NLM supplied data.

The NIH Merit Award was presented to three individuals. The individual recipients were **Loren Frant**, Public Services Division, LO for exceptional contributions to MedlinePlus, MedlinePlus en espanol, and MedlinePlus Mobile; **Stuart J. Nelson, MD**, Medical Subject Headings, LO, for significant leadership in developing and creating

biomedical and drug vocabularies and records systems improving clinical and research communications in the United States; and **Mindy I. Nicolas** for sustained exceptional service to the National Library of Medicine's Extramural Programs.

The NIH Director's Award was presented to one NLM staff member as an individual award and three NLM staff as part of NIH group awards. Individual award was presented to **Suresh Srinivasan**, to recognize his contributions to biomedical terminology research and development efforts at the NIH. The National Eye Institute, Age-Related Eye Disease Study dbGaP Working Group Award included **Michael L. Feolo and Don R. Preuss**, for outstanding coordination and collaboration between NCBI and NEI in expanding dbGaP resources to include the first photographic data on phenotype for age-related macular degeneration. The National Institute of Environmental Health Sciences, Deepwater Horizon Gulf Oil Spill Team Group award included **Perti J. Hakkinen, PhD, SIS**, for the 2010 Gulf Oil Spill response, implementing public health protection programs for cleanup workers, developing the Gulf Study, and assembling a research consortium.

The EEO Special Achievement Award was presented to **George W. Franklin**, office of the NLM Director, for his outstanding contributions in recruitment, training, and organizing the cadre of NLM staff that conduct the NLM Pow-Wow outreach activities.

**Table 13: FY2011 Full-Time Equivalents (Actual)**

<b>Program Area</b>	<b>Count</b>
Office of the Director	9
Office of Health Information Programs Development	5
Office of Communication and Public Liaison	9
Office of Administration	61
Office of Computer and Communications Systems	48
Extramural Programs	16
Lister Hill National Center for Biomedical Communications	66
National Center for Biotechnology Information	275
Specialized Information Services	42
Library Operation	<u>273</u>
<b>TOTAL FTEs</b>	<b>804</b>

**NLM Diversity Council**

In 1998, the NLM Director, Donald A.B. Lindberg, MD, established the Director's Employee Education Fund to empower the increasingly multi-cultural and diverse NLM workforce to enhance and advance their careers and reach their fullest potential. In 2011, the NLM Diversity Council continued the coordination and management of the Director's Education Fund and made continuing education possible for 52 employees to take 83 courses, the majority

of which were undergraduate courses. The majority of NLM staff who took advantage of this opportunity were enrolled in the University of Maryland (32), followed by Strayer University (8) and Montgomery College (5). Other institutions attended included: Catholic University, American University, George Mason University, Georgetown University, Johns Hopkins University, Notre Dame College, Kent State University, Walden University, Carroll Community College, Frederick Community College, University of the District of Columbia, Cappella University, and Columbia Union College. Course disciplines included: Public Health Informatics; Special Topics in Librarianship; Public Programs; Outreach, Management and Organizational Theory; Doctoral Dissertation Seminars, Information Access Services; Marketing Management; Principles and Issues of Information Systems; Constitutional Law; Molecular and General Genetics; Web Design and Development, Interfaces for Information Visualization and Retrieval, among others. The Diversity Council continues to publicize the availability of the fund to NLM employees.

*NLM 2011 Health Expo/Bring Your Child to Work Event*

On April 22nd 2011, the NLM Diversity Council sponsored the 4th annual "Healthy Lifestyles for You and Your Family Expo" in conjunction with NIH's "Take Your Child to Work Day." This was the fourth year for what has now become an annual event. Once again the NLM Diversity Council hosted activities that featured a full day's worth of educational, family-oriented activities focusing on healthy living styles for both adults and children. The Emmy Award-winning FoodPlay Productions which uses both live theatre and interactive media to promote healthy eating and exercise. While FoodPlay makes good eating great fun, its messages are very serious. Childhood obesity has become the nation's number one health epidemic, with rates doubling among elementary school children and tripling among adolescents in the last 25 years. According to the Center for Disease Control and Prevention, if current eating and exercise habits do not improve, one out of two African American and Latino children, and one out of three Caucasian children will develop diabetes. Due to severe weather forecasts, the NLM Diversity Council was forced to shut down all outside activities. Despite these last minute adjustments, NLM was able provide the following sponsored events: Totally Toxic; Children's Book Hospital; Earth Day Jeopardy; The Art of Doll Making Class; and two scientific workshops from Mad Science: Taste, and Smell and Cells. This year's Expo, attended by more than 2,700 NIH and FDA employees and their families, is becoming one of the most popular events held at the NIH. With this overwhelming response, NLM is planning a repeat performance next year.

## Appendix 1: Regional Medical Libraries

1. **MIDDLE ATLANTIC REGION**  
Middle Atlantic Region  
University of Pittsburgh  
Health Sciences Library System  
200 Scaife Hall, 3550 Terrace Street  
Pittsburgh, Pennsylvania 15261  
Phone: (412) 648-2065 Fax: (412) 624-1515  
States served: DE, NJ, NY, PA  
*URL:* <http://nnlm.gov/mar>
2. **SOUTHEASTERN/ATLANTIC REGION**  
University of Maryland at Baltimore  
Health Science and Human Services Library  
601 Lombard Street  
Baltimore, MD 21201-1583  
Phone: (410) 706-2855 Fax (410) 706-0099  
States served: AL, FL, GA, MD, MS, NC,  
SC, TN, VA, WV, DC, VI, PR  
*URL:* <http://nnlm.gov/sea/>
3. **GREATER MIDWEST REGION**  
University of Illinois at Chicago  
Library of the Health Sciences (M/C 763)  
1750 West Polk Street  
Chicago, IL 60612-4330  
Phone: (312) 996-2464 Fax (312) 996-2226  
States served: IA, IL, IN, KY, MI, MN,  
ND, OH, SD, WI  
*URL:* <http://nnlm.gov/gmr>
4. **MIDCONTINENTAL REGION**  
University of Utah  
Spencer S. Eccles Health Sciences Library  
10 North 1900 East  
Salt Lake City, Utah 84112-5890  
Phone: (801) 587-3412 Fax: (801) 581-3632  
States Served: CO, KS, MO, NE, UT, WY  
*URL:* <http://nnlm.gov/mcr>
5. **SOUTH CENTRAL REGION**  
Houston Academy of Medicine-  
Texas Medical Center Library  
1133 MD Anderson Boulevard  
Houston, TX 77030-2809  
Phone: (713) 799-7880 Fax: (713) 790-7030  
States served: AR, LA, NM, OK, TX  
*URL:* <http://nnlm.gov/scr>
6. **PACIFIC NORTHWEST REGION**  
University of Washington  
Health Sciences Libraries and  
Information Center  
Box 357155  
Seattle, WA 98195-7155  
Phone: (206) 543-8262 Fax: (206) 543-2469  
States served: AK, ID, MT, OR, WA  
*URL:* <http://nnlm.gov/pnr>
7. **PACIFIC SOUTHWEST REGION**  
University of California, Los Angeles  
Louise M. Darling Biomedical Library  
Box 951798  
Los Angeles, CA 90025-1798  
Phone: (310) 825-1200 Fax: (310) 825-5389  
States served: AZ, CA, HI, NV and  
US Territories in the Pacific Basin  
*URL:* <http://nnlm.gov/psr>
8. **NEW ENGLAND REGION**  
University of Massachusetts Medical School  
222 Maple Avenue  
Shrewsbury, MA 01545  
Phone: (508) 856-5979 Fax: (508) 856-5977  
States Served: CT, MA, ME, NH, RI, VT  
*URL:* <http://nnlm.gov/ner>

## Appendix 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.

TANJI, Virginia, MSLS (**Chair**)  
Director  
Health Science Library  
University of Hawaii at Manoa  
School of Medicine  
Honolulu, HI

EVENS, Ronald G., MD  
Professor  
Mallinckrodt of Radiology  
Washington University School of Medicine  
Saint Louis, MO

FLEMING, David A., MD  
UAMS Library Director/Professor  
Department of Internal Medicine  
University of Missouri School of Medicine  
Columbia, MO

GOTTLIEB, Katherine, PhD  
President and CEO  
Southcentral Foundation  
Anchorage, AK

LEWIS, Henry, PharmD  
President  
Florida Memorial University  
Miami Gardens, FL

MACKAY, Trudy, PhD  
Distinguished University Professor of Genetics  
Department of Genetics  
North Carolina State University  
Raleigh, NC

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

ROSKIES, Ralph Z., PhD  
Professor of Physics and Scientific Director  
Department of Physics and Astronomy  
Pittsburgh Supercomputing Center  
University of Pittsburgh  
Pittsburgh, PA 15213

RYAN, Mary L., MLS, MPH, BA  
UMAS Library Director/Professor  
University of Arkansas for Medical Sciences Library  
Little Rock, AR

SCUTCHFIELD, F. Douglas  
Peter P Bosomworth Professor of Health Services  
Research and Policy  
University of Kentucky College of Public Health  
Lexington, KY

### EX OFFICIO MEMBERS

LIU, Simon Y., PhD  
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

GREEN, Charles B., 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

PETZEL, Robert, MD  
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.

ASH, Joan S., PhD (**Chair**)  
Professor and Vice-Chair  
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

CHAPMAN, Wendy W., PhD  
Associate Professor  
Division of Biomedical Informatics  
University of California, San Diego  
La Jolla, CA

CHUEH, Henry C., MD  
Director and Chief  
Laboratory of Computer Science  
Division of Biomedical Informatics

Massachusetts General Hospital  
Boston, MA

HAMMOND, Steven C., MD  
Professor of Radiology and Clinical Director  
Medical Informatics Group  
Department of Radiology  
Hospital of the University of Pennsylvania  
Philadelphia, PA

MANDL, Kenneth D., MD  
Associate Professor  
Harvard Medical School  
Director, Intelligent Health Laboratory  
Children's Hospital Informatics Program  
Children's Hospital Boston  
Boston, MA

WERE, Martin C., MD  
Assistant Professor of Medicine  
Department of General Internal Medicine and Geriatrics  
Indiana University School of Medicine  
Indianapolis, IN

## 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.

LEVINE, Arthur S., MD (**Chair**)  
Senior Vice Chancellor for Health Sciences  
Dean, School of Medicine  
University of Pittsburgh  
Pittsburgh, PA

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

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

SEIDMAN, Christine E., MD  
T.W. Smith Professor of Medicine and Genetics  
Harvard Medical School  
Boston, MA

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

COOPER, Gregory F., MD, PhD (**Chair**)  
Professor  
Department of Biomedical Informatics  
University of Pittsburgh  
Pittsburgh, PA

ALPI, Kristine M., MLS, MPH  
Director & Adjunct Professor  
Department of Population Health & Pharmacology  
Veterinary Medical Library  
North Carolina State University  
Raleigh, NC

CARDOZO, Timothy J., MD, PhD  
Associate Professor Department of Biochemistry and  
Molecular Pharmacology  
New York University School of Medicine  
New York, NY

GONAZALEZ, Graciela H., PhD  
Associate Professor  
Department of Biomedical Informatics  
Mayo Clinic  
Arizona State University  
Indianapolis, IN

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

KAHN, Michael G., MD, PhD  
Associate Professor  
Department of Pediatrics  
University of Colorado  
Aurora, CO

KALET, Ira J., PhD  
Professor Emeritus  
Radiation Oncology Department  
Department of Biomedical Informatics and Medical  
Education  
University of Washington  
Seattle, WA

KOCHI, Julia K., MLS  
Director, Digital Library & Collections  
Library & Center for Knowledge Management  
University of California, San Francisco  
San Francisco, CA

LUSSIER, Yves. A, MD  
Professor of Medicine and Engineering  
Clinical Research Information Office  
Assistant Vice President of Health Affairs  
The University of Illinois in Chicago  
Chicago, IL

MOORE, Jason H., PhD  
Professor of Genetics  
Dartmouth-Hitchcock Medical Center  
Dartmouth College  
Lebanon, NH

MURPHY, Beverly, MLS  
Assistant Director  
Marketing and Publication  
Webmaster, DUMCL Online  
Duke University Medical Center Library  
Durham, NC

PESTIAN, John P., PhD  
Professor  
Director, Computational Medicine Center  
Cincinnati Children's Hospital Medical Center  
University of Cincinnati  
Cincinnati, OH

SHATKAY, Hagit, PhD  
Associate Professor  
Department of Computer and Information Sciences  
University of Delaware  
Newark, DE

SITTIG, Dean F., PhD  
Professor  
UT Memorial Hermann Center for Healthcare Quality and  
Safety  
The University of Texas School of Health Information  
Sciences at Houston  
Houston, TX

SRINIVASAN, Padmini, PhD  
Professor  
Computer Science  
The University of Iowa  
Iowa City, IA

STARREN, Justin, MD, PhD  
Chief, Division of Biomedical Informatics, Department of  
Preventive Medicine  
Director, Northwestern University Biomedical Informatics  
Center  
Feinberg School of Medicine, Northwestern University  
Chicago, IL

SUBRAMANIAM, Shankar, PhD  
Professor and Chair  
Department of Bioengineering  
University of California, San Diego  
La Jolla, CA

VEINOT, Tiffany C.E., PhD  
Assistant Professor  
School of Information  
University of Michigan  
Ann Arbor, MI

## 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.

CHRISTOPHER, Mary M., PhD (**Chair**)  
Professor of Pathology  
Dept. of Pathology, Microbiology & Immunology  
School of Veterinary Medicine  
University of California  
Davis, CA

CABELLO, Felipe C., MD  
Professor  
Dept. Of Pathology, Microbiology & Immunology  
New York Medical College  
Valhalla, NY

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

COURTNEY, Karen L., PhD  
Associate Professor  
School of Health Information Science  
University of Victoria  
Victoria, BC

CRUMMETT, Courtney, MS, MLS  
Bioinformatics Librarian  
Massachusetts Institute of Technology  
Cambridge, MA

GWINN, Marta, MD  
Senior Consultant  
Office of Public Health Economics  
Centers for Disease Control and Prevention  
Atlanta, GA

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

OGUNYEMI, Omolola, PhD  
Director  
Center for Biomedical Informatics  
Charles Drew University of Medicine and Science  
Lynwood, CA

PASCOE, John M., MD  
Professor  
Department of Pediatrics  
Wright State University  
Dayton, OH

PHILLIPS, William R., MD  
Clinical Professor  
Department of Family Medicine  
University of Washington  
Seattle, WA

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

## 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

ANDERSON, Ivy, MLS  
Director  
Collection Development & Management Program  
California Digital Library  
Oakland, CA

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

BOURNE, Philip E., PhD  
Professor  
Department of Pharmacology  
San Diego Supercomputer Center  
University of California, San Diego  
La Jolla, CA

COLAMARINO, Sophia A., PhD  
Consulting Associate Professor  
Department of Psychiatry and Behavioral Sciences  
Stanford University Medical School  
San Francisco, CA

COURANT, Paul N., PhD  
University Librarian and Dean of Libraries  
University of Michigan Library  
University of Michigan  
Ann Arbor, MI

FASSLER, Jan S., PhD  
Professor  
Department of Biology  
University of Iowa  
Iowa City, IA

HENDERSON, Cynthia L., MLS  
Executive Director  
Louis Stokes Health Science Library  
Howard University  
Washington, DC

KANN, Maricel G., PhD  
Assistant Professor  
Department of Biological Sciences  
University of Maryland  
Baltimore, MD

MEGLIO, Delores, MS  
Vice President, Public Relations  
Knovel Corporation  
New York, NY

ROSSNER, Mike, PhD  
Executive Director  
The Rockefeller University Press  
New York, NY

TANNER, R. Michael., PhD  
Vice President and Chief Academic Officer  
Association of Public and Land-Grant Universities  
Washington, DC

THIBODEAU, Patricia L., MLS, MBA  
Associate Dean for Library Services & Archives  
Medical Center Library  
Duke University  
Durham, NC

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

## Appendix 8: Organizational Acronyms and Initialisms Used in this Report

<u>Acronym</u>	<u>Meaning of Acronym</u>	<u>Acronym</u>	<u>Meaning of Acronym</u>
AAHSL	Association of Academic Health Sciences Libraries	BoSC	Board of Scientific Counselors
AABB	Non-profit association formerly known as American Association of Blood Banks	BSD	Bibliographic Services Division
AAPA	American Academy of Physicians Assistants	BSN	Bioinformatics Support Network
ABC	Advanced Biomedical Tele-Collaboration Test Bed)	CAM	Complementary and Alternative Medicine
ACLA	American Clinical Laboratory Association	C&A	Certification & Accreditation (audit)
ACORN	Automatically Creating OLDMEDLINE Records for NLM	CANDHI	Central American Network for Disaster and Health Information
ACP	American College of Physicians	CARE	Candidate Gene Association Resource project
ACSI	American Customer Satisfaction Index	CAS	Collection Access Section
AFIP	Armed Forces Institute of Pathology	CBB	Computational Biology Branch
AG	Access Grid	CBIR	Content-Based Image Retrieval
AHIC	American Health Information Community	CCB	Configuration Control Board
AHRQ	Agency for Healthcare Research and Quality	CCDS	Consensus CoDing Sequence
AIDS <i>info</i>	Acquired Immune Deficiency Syndrome <i>info</i> (database)	CCHIT	Commission for Healthcare Information Technology
ALTBIB	Alternatives to Animal Testing	CCR	Central Contractor Registration
AME	Automated Metadata Extraction	CCRIS	Chemical Carcinogenesis Research Information System
AMIA	American Medical Informatics Association	CDD	Conserved Domain Database
AMPA	American Medical Publishers Association	cDNA	Complementary DNA
AMPATH	Academic Model Providing Access to Healthcare	CEB	Communications Engineering Branch
AMWA	American Medical Women's Association	CEL	Affymetrix Cell intensity (file)
APDB	Audiovisual Program Development Branch	CgSB	Cognitive Science Branch
API	Applied Programming Interface	CHEBI	Chemical Entities of Biological Interest
APIRE	American Psychiatric Institute for Research and Education	ChEMBL	Computational Chemical Biology Group database
ARRA	American Recovery and Reinvestment Act	ChemIDplus	Chemical Identification File
ASCCP	American Society for Cervical Pathology and Colposcopy	CHEMM	Chemical Hazard Event Medical Management
ASHG	American Society of Human Genetics	CHRIS	Consumer Health Resource Information Service
ASPR	Assistant Secretary for Preparedness and Response, HHS Office of the	CHIC	Chickasaw Health Information Center
BAC	Bacterial Artificial Chromosome	CIT	Center for Information Technology
BarSTool	Barcode Submission Tool	CLML	Current List of Medical Literature
BGMUT	Blood Group Antigen Gene Mutation Database	CMAX	Collaborative Multi-Agency eXercise (BHEPP disaster drill)
BHEPP	Bethesda Hospitals' Emergency Preparedness Partnership	CMS	Centers for Medicare and Medicaid Services
BISTI	Biomedical Information Science and Technology Initiative	CMT	Convergent Medical Terminology
BITA	Biomedical Image Transmission via Advanced Networks	COOP	(NIH Pandemic Flu) Continuity of Operations Plan
BLAST	Basic Local Alignment Search Tool	CORE	Clinical Observations Recording and Encoding
BLIRC	Biomedical Library and Informatics Review Committee	CoreBio	Core Bioinformatics Facility
BMT	Boundary Marking Tool	CPS	Commercial Peering Service
BN	Brand Name	CPSC	Center for Public Service Communication
BOR	Board of Regents	CPT	Current Procedural Terminology
BSAT	BMT Study Administration Tool	CRAC	Computer Room Air Conditioner

<u>Acronym</u>	<u>Meaning of Acronym</u>
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
CTD	Clinical Text De-identification
CTD	Comparative Toxicogenomics Database
CTS	Communications Technology Satellite
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
DBA	Data Base Administrator
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
DCMS	Data Creation and Maintenance System
DDBJ	DNA Data Bank of Japan
DDD	Drug Delivery Devices
DDoS	Distributed Denial of Service (attack)
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
EAI	Emergency Access Initiative
EBI	European Bioinformatics Institute
EBP	Evidence-Based Practice
ECHO	European Community Humanitarian Office
Educollab	Educational Collaborators
EEO	Equal Employment Opportunity
EFTS	Electronic Funds Transfer Service
EHR	Electronic Health Record

<u>Acronym</u>	<u>Meaning of Acronym</u>
EMBL	European Molecular Biology Laboratory
EMR	Electronic Medical Record EMS Emergency Medical Services
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
eTK	Electronic Thorndike and Kibre
EUREKA	Exceptional, Unconventional Research Enabling Knowledge Acceleration
E-Utilities	Entrez Programming Utilities
eVK	Electronic Voights and Kurtz
FAES	Foundation for Advanced Education in the Sciences
FDA	Food and Drug Administration
FDCC	Federal Desktop Core Configuration
FHA	Federal Health Architecture
FIC	Fogarty International Center
FISMA	Federal Information Security Management Act
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
GDP	Genome Decoration Page
GDS	GEO DataSet
GEO	Gene Expression Omnibus (database)
GENSAT	Gene Expression Nervous System Atlas
geneRIF	Gene Reference Into Function
GENE-TOX	Genetic Toxicology
GHR	Genetics Home Reference
GIS	Geographic Information System
GO	Grand Opportunity grant
GO-ESP	Grand Opportunities-Exon Sequencing Project
GMAC	Grants Management Advisory Committee
GPS	Global Position System
GPU	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

<u>Acronym</u>	<u>Meaning of Acronym</u>
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
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
ICC	Incident Command Center
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
IGS	Intergenic Spacer
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)
infoSIDA	infoSíndrome de Inmunodeficiencia Adquirida (database)

<u>Acronym</u>	<u>Meaning of Acronym</u>
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
ITP	Informatics Training Program
ITER	International Toxicity Estimates for Risk
ITK	Insight Toolkit
ITP	Informatics Training Program
ITS	Internal Transcribed Space
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
LC	Library of Congress
LHNCBC	Lister Hill National Center for Biomedical Communications
LID	Laboratory for Informatics Development
LITE	Librarian Infobutton Tailoring Environment
LJI	List of Journals Indexed
LO	Library Operations
LOINC	Logical Observations Identifiers, Names, Codes
LPF	Lost Person Finder
LRP	Long Range Plan (NLM)
LSD	Lysosomal Storage Disorders
LSI	List of Serials Indexed
LSTRC	Literature Selection Technical Review Committee
LVG	Lexical Variant Generator
LWS	Lifecycle Work Station
MARC	Machine-Readable Cataloging
MARG	Medical Article Records Groundtruth
MARS	Medical Article Records System
MAX	Mid Atlantic Exchange, U. of Maryland
MCI	Mass Casualty Incident
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

<u>Acronym</u>	<u>Meaning of Acronym</u>
MEO	Medical Education and Outreach
MeSH	Medical Subject Headings
MHC	Major Histocompatibility Complex
MHL	Medical Heritage Library
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
mRNA	Messenger Ribonucleic Acid
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
NCBC	National Center for Biomedical Computing
NCBI	National Center for Biotechnology Information
NCCS	NIH Consolidated Collocation Site
NCHS	National Center for Health Statistics
NCMHD	National Center for Minority Health and Health Disparities
NCI	National Cancer Institute
NCRR	National Center for Research Resources
NCVHS	National Committee on Vital and Health Statistics
NDC	National Data Codes
N <sub>e</sub> 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

<u>Acronym</u>	<u>Meaning of Acronym</u>
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
NIHES	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
NLM LitArch	NLM Literature Archive
NLP	National Language Processing System
NN/LM	National Network of Libraries of Medicine
NNMC	National Naval Medical Center
NNO	National Network Office
NOAA	National Oceanic and Atmospheric Administration
NOSC	Network Operations and Security Center
NOVA	National Online Volumetric Archive
NPL	National Priorities List (as for Superfund)
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
OCIO	Office of the Chief Information Officer (NIH)
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

<u>Acronym</u>	<u>Meaning of Acronym</u>	<u>Acronym</u>	<u>Meaning of Acronym</u>
OMD	Office of Management and Budget	RCSB	Research Collaboratory for Structural Bioinformatics
OMIA	Online Inheritance in Animals (database)	RDMS	Rare Disease Maintenance System
OMIM	Online Mendelian Inheritance in Man (database)	RefSeq	Reference Sequence (database)
OMSSA	Open Mass Spectrometry Search Algorithm	REMIM	Radiation Even Medical Management
ONC	Office of National Coordinator (for Health Information Technology)	RF2	Release Format version 2
OPASI	Office of Portfolio Analysis and Strategic Initiatives	RFA	Request for Applications
OPD	OLDMEDLINE Serials Application	RFID	Radio Frequency Identification
OPD	Outreach Products Database	RFP	Request for Proposals
OSA	Optical Society of America	RHIN	Refugee Health Information Network
ORWH	Office of Research on Women's Health	RIDeM	Repository for Informed Decision Making
OSIRIS	Open Source Independent Review and Interpretation System	RML	Regional Medical Library
PAHO	Pan American Health Organization	RNA	Ribonucleic Acid
PAL	Potential Abstract Labels (tool)	RNAi	RNA Interference
PBM	Pharmacy Benefit Manager	RPS-BLAST	Reversed Position Specific BLAST
PCA	Personal Computer Advisory Committee	RQS	Request Submission and Tracking System
PCR	Polymerase Chain Reaction	RRF	Rich Release Format
PDA	Personal Digital Assistant	RSS	Really Simple Syndication
PDRS	Publisher Data Review System	RTECS	Registry of Toxic Effects of Chemical Substances
PDB	Protein Data Bank	RTLS	Real Time Location System
PDF	Portable Document Format	RVDS	Remote Virtual Dialogue System
PDL	Personal Digital Library	RWJF	Robert Wood Johnson Foundation
PDM	Patient Data Management	SAB	Source Abbreviations
PFIF	Person Finder Interchange Format	SBIR	Small Business Innovation Research
PheGenI	Phenotype-Genotype Integrator	SCID	Severe Combined Immunodeficiency
PHLIP	Public Health Law Information Project	SCR	(MeSH) Supplemental Chemical Records
PHII	Public Health Informatics Institute	SDK	Software Development Kit
PHP	Public Health Partners	SEF	Serials Extract File
PHR	Personal Health Record	SEIM	Security Event and Information Management System
PHS	Public Health Service	SEO	Search Engine Optimization
PI	Pathway to Independence award	SEP	Special Emphasis Panel
PI	Principal Investigator	SEQ	Structured Evidence Queries
PII	Personally Identifiable Information	SIDA	Swedish International Development Agency
PICO	Patient/Population, Intervention, Comparison, and Outcome	SIG	Special Interest Group
PID	Pathway Interaction Database (NCI)	SII	Scalable Information Infrastructure
PIN	Precise Ingredient (term type), RxNorm	SIS	Specialized Information Services
PL	Person Locator	SKR	Semantic Knowledge Representation
PLAWARe	Programmable Layered Architecture With Artistic Rendering	SMART	Scalable Medical Alert and Response Technology
PMC	PubMed Central	SNOMEDCT	Systematized Nomenclature of Medicine Clinical Terms
PMCI	PubMed Central International	SO	Signing Official
PMC ID	PubMed Central Identification (number)	SOAP	Simple Object Oriented Protocol (formerly Simple Object Access Protocol)
PRS	Protocol Registration System	SPER	System for the Preservation of Electronic Resources
PSD	Public Services Division	SPIN	Shared Pathology Informatics Network
PubMedHh	PubMed for Handhelds	SPIRS	Spine Pathology Image Retrieval System
PUG	PubChem Power User Gateway	SRA	Short Read Archive
QA	Quality Assurance	STB	Systems Technology Branch
QCIM	Quarterly Cumulative Index Medicus	STEM	Science, Technology, Engineering and Math
RAC	Real Application Clusters	STR	Short Tandem Repeat
RCDC	Research Condition and Disease Categorization	STTP	Short-Term Trainee Program

<u>Acronym</u>	<u>Meaning of Acronym</u>
STTR	Small Business Technology Transfer Research
STS	Sequence Tagged Site
SVM	Support Vector Machine
TA	Title Abbreviation(s)
TBL	The bottom line
TDI	3D Informatics (Group)
TEHIP	Toxicology and Environmental Health Information Program
TERA	Toxicology Excellence for Risk Assessment
TIC	Trusted Internet Connection
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

<u>Acronym</u>	<u>Meaning of Acronym</u>
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)
UIMA	Unstructured Information Management Architecture
UKPMC	United Kingdom PubMed Central
UMLS	Unified Medical Language System
UMLSKS	UMLS Knowledge Source Server
UN	United Nations

*Further information about the programs described in this  
Administrative report is available from:*

*Office of Communications and Public Liaison*

*National Library of Medicine*

*8600 Rockville Pike*

*Bethesda, MD 20894*

*301-496-6308*

*E-mail: [publicinfo@nlm.nih.gov](mailto:publicinfo@nlm.nih.gov)*

*Web: [www.nlm.nih.gov](http://www.nlm.nih.gov)*

**Cover:**

The National Library of Medicine marked its 175th anniversary in 2011. As part of the celebration, hundreds of workers formed the number 175 and posed for this commemorative photo. Much has changed since 1836. What began as a small collection of books in the Office of the Army Surgeon General has grown to become the world's largest medical library and the worldwide leader in trusted health information and innovation. *Photo by Jessica Marcotte, 2011.*

