NATIONAL INSTITUTES OF HEALTH

National Library of Medicine

Programs and Services

Fiscal Year 2013

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

Preface ..... **v**

Office of Health Information Programs Development **1**
- Planning and Analysis 1
- Outreach and Consumer Health 1
- International Programs 3
- Trans-NIH Collaborations in Data Science 4

Library Operations **5**
- Program Planning and Management 5
- Collection Development and Management 5
- Preservation and Collection Management 6
- Vocabulary Development and Standards 7
- Bibliographic Control 9
- Information Products 10
- Direct User Services 12
- Outreach and Training 12

Specialized Information Services **21**
- Toxicology and Environmental Health Resources 22
- Disaster Information 24
- Outreach Initiatives 27
- Social Media 29
- K-12 Initiatives and Resources 29
- HIV/AIDS 30
- Evaluation Initiatives 31

Lister Hill National Center for Biomedical Communications **34**
- Next Generation Electronic Health Records to Facilitate Patient-centric Care, Clinical Research, and Public Health 34
- Biomedical Imaging, Multimedia, and 3D Imaging 36
- Natural Language Processing and Text Mining 42
- Information Resource Delivery for Researchers, Care Providers, and the Public 46
- Clinical Vocabulary Standards and Associated Tools 48
- Communication Infrastructure Research and Tools 51
- Disaster Information Management: Lost Person Finder 52
- Video Production, Retrieval, and Reuse Project 53
- Computing Resources Projects 54
- Training and Education at LHNCBC 55

National Center for Biotechnology Information **56**
- Molecular Biology Information Resources 56
- dbGaP: Integration of Clinical and Genetic Data 65
- Entrez Retrieval System 65
- Literature Information Resources 66
- Research 67
- Bioinformatics Training and Support 68

Extramural Programs **70**
- Overview of FY2013 70
- Research Grants 75
- Resource Grants 75
- Training and Career Awards 75
- Early Recruitment to Informatics Research Careers 76
- Trans-NIH Initiatives 77

Pan-NIH Projects and Interagency Collaborations 77
Extramural Programs Web Site 78
Scientific Review 78

Office of Ccomputer and Communications Systems **85**
Controlled Medical Vocabularies 85
Consumer and Public Health 86
IT Infrastructure Services 88
Medical Literature Support & Document Delivery Services 93
Outreach and Customer Services 94

Administration **96**
Financial Resources and Allocations, FY2013 96
Personnel 96
NLM Diversity Council 100
NLM 2013 Health Expo/Take Your Child to Work Event 100

NLM Organization Chart **117**

Appendixes

1. RFA/PA Actions in FY2013 ..... **79**
2. NLM New Grants Awarded in FY2013 ..... **81**
3. FY2013 Grant Budget by NIH Mechanism and NIH Activity Code **83**
4. Regional Medical Libraries **101**
5. Board of Regents **102**
6. Board of Scientific Counselors, Lister Hill Center for Biomedical Communications **103**
7. Board of Scientific Counselors, National Center for Biotechnology Information **104**
8. Biomedical Library and Informatics Review Committee **105**
9. Literature Selection Technical Review Committee **107**
10. PubMed Central National Advisory Committee **108**
11. Organizational Acronyms and Initialisms Used in this Report **109**

Figures

Figure 1. FY2013 Sequestration Effects on New Research Project Grants **70**
Figure 2. FY2013 EP Grants Budget **71**
Figure 3. Publications from Active Grants in FY2013 ..... **73**
Figure 4. Top 10 Publications where NLM Grantees Published, 2013 ..... **74**
Figure 5. Virtualization Program Progress **91**
Figure 6. Growth of Enterprise Storage **92**

Tables

Table 1. Collections **16**
Table 2. Collection Activities **17**
Table 3. Cataloging and Indexing **18**
Table 4. Services to the Public **19**
Table 5. Selected Web Resources **20**
Table 6. FY2013 Active Projects by Investment Area **71**
Table 7. Financial Resources and Allocations, FY2013 ..... **96**
Table 8. FY2013 Full-Time Equivalents (Actual) **96**

PREFACE

In Fiscal Year 2013, the National Library of Medicine again saw a wide range of accomplishments in many areas. Notable achievements included the following:

- Holding to our tradition of responsiveness to disasters both manmade and natural, our Disaster Information Management Research Center (DIMRC) launched a special page providing extensive health resources related to Hurricane Sandy. The site includes links to overviews, state-specific pages, cleanup and recovery information, mental health information, multi-language resources, social media information, apps and widgets, and more.

- Genetics Home Reference celebrated 10 years of public service. The consumer-friendly site provides a bridge between the public's questions about human genetics and the rich technical data that has emerged from the Human Genome Project (observing the 10th anniversary of its completion on April 25th, the same day of GHR's) and other genomic research.

- Now approaching its 25th anniversary, the National Center for Biotechnology Information (NCBI) continues to astonish us with the quality and variety of its resources, and the frequency of their use. MEDLINE/PubMed saw approximately 2.5 billion searches in FY2013, a 14 percent increase from last year. Page views in PubMed totaled 7.57 billion. MEDLINE/PubMed now includes more than 23 million citations.

- Over 298 million visitors viewed over 845 million pages on the MedlinePlus full Web site. MedlinePlus Mobile received over one million visitors viewing over 4.5 million pages. MedlinePlus remains in the top five of government sites based on the American Customer Satisfaction Index (ACSI).

- In partnership with the NIH Office of Dietary Supplements, we launched the Dietary Supplement Label Database (DSLD). Researchers, as well as health care providers and consumers, can now see the ingredients listed on the labels of about 17,000 dietary supplements, and the information can be searched and organized as desired.

- In collaboration with the Office of the National Coordinator for Health Information Technology (ONC) and Centers for Medicare & Medicaid Services (CMS), we launched the NLM Value Set Authority Center (VSAC). Initially the VSAC will provide downloadable access to all official versions of vocabulary value sets contained in the 2014 Clinical Quality Measures. Content will gradually expand to incorporate value sets for other use cases, as well as for new measures and updates to existing measures.

- We released about 1,800 new high quality images of solid oral dosage medications. Researchers and product developers can obtain the images and accompanying metadata free of charge via an applications programming interface (API), and they are also available for interactive Web searching via our Pillbox and RxNav sites.

- The History of Medicine Division launched a lively new blog, *Circulating Now*, to encourage greater exploration and discovery of one of the world's largest and most treasured history of medicine collections. HMD also announced the loan of several items to the collection to the Grolier Club, the Metropolitan Museum of Art and the Smithsonian's Sackler Gallery.

I commend our talented and committed staff, advisors and consultants for continuing to hew to NLM's high standard of excellence and innovation, always with an eye to public service and improving the public health.

Donald A.B. Lindberg, MD
Director

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Michael F. Huerta, PhD
Associate Director

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

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

Planning and Analysis

The NLM Long Range Plan remains at the heart of NLM's planning and budget activities. Its goals form the basis for NLM operating budgets each year. *Charting a Course for the 21st Century: NLM's Long Range Plan 2006–2016* is available in print and on the NLM Web site. Print copies are available from the NLM Office of Communications and Public Liaison. The report is organized around four key goals:

- Goal 1. Seamless, Uninterrupted Access to Expanding Collections of Biomedical Data, Medical Knowledge, and Health Information
- Goal 2. Trusted Information Services that Promote Health Literacy and the Reduction of Health Disparities Worldwide
- Goal 3. Integrated Biomedical, Clinical, and Public Health Information Systems that Promote Scientific Discovery and Speed the Translation of Research into Practice
- Goal 4. A Strong and Diverse Workforce for Biomedical Informatics Research, Systems Development, and Innovative Service Delivery

OHIPD also has overall responsibility for developing and coordinating the NLM Health Disparities Strategic Plan, outlining NLM strategies and activities undertaken in support of NIH efforts to understand and eliminate health disparities between minority and majority populations. This plan addresses priorities related to improving access to health information as a means to address health disparities, with special emphasis on rural, minority, and other underserved populations. Areas of emphasis include programs in research and research capacity building as well as information dissemination, community outreach, and public health education. NLM's Health Disparities Strategic Plan for 2009-2013 is available on the NLM Web site.

Outreach and Consumer Health

NLM carries out a diverse set of activities directed at building awareness and use of its products and services by health professionals and by the general public.. Considerable emphasis is placed on reducing health disparities by targeting health professionals and consumers in frontier, rural and inner city areas. NLM's outreach projects build on long experience with addressing the needs of health professionals and making consumers aware of health information resources, particularly in the HIV/AIDS area and for senior citizens, Native American communities, and the Spanish-speaking public. In FY2013, OHIPD staff continued outreach initiatives intended to encourage underrepresented minority high school students to utilize NLM's health information resources and to pursue careers in medicine and the health sciences, carried out in collaboration with other divisions of NLM. An NLM-wide Coordinating Committee on Outreach, Consumer Health and Health Disparities (OCHD) plans, develops, and coordinates NLM outreach and consumer health activities. The OCHD is chaired and staffed by OHIPD.

Support for the Mentoring In Medicine (MIM) Science and Health Career Exploration program continued in FY2013. This initiative reaches seven public and charter schools with an after school program to enrich the high school biology curriculum and encourage enrollment in higher education programs leading to degrees in medicine, allied health professions, and medical librarianship. Principals, science teachers and guidance counselors from participating schools oversee 40 sessions of biology instruction in 12 organ systems, taught by visiting health professionals/mentors over a two-year period. During its five-year operation, the program has exposed over 800 minority students to health care career instruction. In the process, the program has developed four after-school curricula, nine knowledge tests and a series of pre and post course evaluation surveys and instructor feedback forms. Program evaluations demonstrate continued impressive gains in health care knowledge. Through a combination of personal intervention and online education resources, the program employs an innovative educational curriculum that also strengthens high school students' readiness to pursue health careers. NLM staff also participated in MIM's 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.

NLM also continues its focused efforts to meet the health information needs of the Hispanic population in Texas and elsewhere. In FY2013, NLM continued its support for the South Texas High School for the Health Professions, known as MedHigh, a magnet health high school in the Lower Rio Grand Valley of Texas. The MedHigh iVIVA! Peer Tutors Program is an award winning effort to involve high schools students in teaching their peers about online health information. The peer tutors also conduct outreach to the local community and sponsor a comprehensive program Web site for 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 students, teachers, and guidance counselors. In FY2013 a photo voice project was conducted to encourage peer tutors to develop vignettes with photos illustrating the impact of using Medlineplus.gov. Peer tutoring is offered at other magnet high schools in the Lower Rio Grande Valley, where it continues to be an effective outreach program. A practical "how to" implementation guide was finalized, to assist other high schools in considering and implementing a peer tutoring program. The guide is based in part on prior research and reports, including a publication addressing the long-term outcomes of the project. (Olney CA, Hansen L, Vickman A, Reibman S, Wood FB, Siegel E. *J. Med. Libr. Assoc.* 2011 Oct; 99 (4): 317-20.)

The information guide and supporting documents were posted on the iVIVA! Web site, http://viva.stisd.net/implementation.html , and also linked to the NLM Specialized Information Systems (SIS) Web site, http://sis.nlm.nih.gov/outreach/vivaproject.html .

Native American Outreach

In FY2013 OHIPD again participated in the NIH American Indian Pow-Wow Initiative to demonstrate the range of NLM information resources for consumer audiences and to enhance awareness of the resources. This included exhibiting at a dozen pow-wows, mostly in the Mid-Atlantic area. An estimated 10,000 persons visited the NLM booth over the course of these pow-wows. These activities are yet another way to increase awareness of NLM's health information resources among segments of the Native American community and the general public.

OHIPD supported (via the Greater Midwest Regional Medical Library) a project at Cankdeska Cikana Community College, Spirit Lake Nation, Ft. Totten, North Dakota that resulted largely from the Native American Listening Circles conducted in prior years and continued into FY2013. This continuing effort supports improvements at the tribal college library and educational and community outreach activities run by the library.

In FY2013, OHIPD supported an upgrade of the Native Health Database maintained by the University of New Mexico Health Sciences Library and Informatics Center.

OHIPD also partnered with Mentoring in Medicine (MIM) and the Montana State University (MSU) School of Nursing on two tribal outreach projects in Montana. In April 2013, NLM, MIM, and MSU hosted a Health Careers Week at the Hardin High School, Hardin, MT. About 70 percent of the students are American Indian, mostly from the Crow Tribe and some from the Northern Cheyenne Tribe. Several hundred students participated in one or more program elements, that included a *Native Voices* Exhibition iPad experience, an in-class health science and pig heart dissection experience, a health careers fair with over a dozen exhibitors, and a general assembly with special speakers. In September 2013, NLM, MIM, and MSU hosted a Health Careers and Healthy Living Seminar at the Little Big Horn College, Crow Agency, MT, a tribal college serving the Crow Indian Tribe. About 60 college students attended an afternoon seminar that included speakers from NLM, local health care providers, and local diabetes prevention educators.

Native Voices Exhibition

In FY2013, OHIPD supported or facilitated *Native Voices* follow-up outreach projects in the Greater Midwest RML with the: Sitting Bull College, Standing Rock Sioux Nation, Ft. Yates, ND; White Earth Tribal and Community College, White Earth Chippewa (Ojibwe) Indian Nation, Mahnomen, MN; and Indian Center of Chicago; and in the Pacific Northwest RML with the: Northwest Indian College, Lummi Indian Tribe, Bellingham, WA; and the Seattle Indian Health Board, Seattle, WA; and with Papa Ola Lokahi, Honolulu, HI, a Native Hawaiian Health Organization; and the Waianae Coast Comprehensive Health Center and Traditional Healing Center, Waiane, Oahu, HI.

OHIPD also assisted with the *Native Voices* traveling exhibition pilot test planning and implementation. Pilot test sites are planned for: Cankdeska Cikana Community college, Spirit Lake Dakota Nation, Ft. Totten, ND; National Congress of American Indians Mid-Year Conference, Alaska Native Heritage Center, and Primary Care Center Southcentral Foundation, all in Anchorage, AK; Queens Medical Center, Honolulu, HI; and, tentatively, the Chickasaw Nation, Sulphur and Ada, OK.

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. The Web evaluation priorities of the OCHD include both quantitative and qualitative metrics of Web usage,

measures of customer perception, and use of NLM Web sites. During FY2013, the OCHD continued to pursue an integrated approach intended to encourage exchange of information and learning within NLM, and help better inform NLM management decision-making on Web site research, development, and implementation. The year's evaluation activities included analysis of NLM Web site log data; continuation a of trans-NLM Web metrics program; and access to Internet audience measurement estimates based on Web usage by user panels organized by a private sector company. During FY2013, OHIPD continued to coordinate NLM's use of the online Web user survey known as the American Customer Satisfaction Index (ACSI). The ACSI provides ongoing user feedback to NLM's Web site manager. OHIPD also coordinated NLM's use of Internet measurement services provided by comScore Inc. via Iron Bow LLC.

OHIPD staff participated in FY2013 in various Web Metrics and Web Analytics seminars and webinars on Web evaluation opportunities and challenges. Many of these learning experiences are open to all interested NLM staff.

International Programs

NLM's international partnerships and projects strengthen and expand global access to the world's health literature. Programs are intended to strengthen all phases of the research process, with particular emphasis by OHIPD on outreach to researchers, physicians, medical students, health workers, journal editors, and librarians in developing countries. Initiatives include demonstration projects that relate to NLM's major programs and databases, including information dissemination strategies and training opportunities. NLM also contributes to the Fogarty International Center's Medical Education Partnership Initiative in Africa (MEPI) program, which provides grants to institutions in Sub-Saharan African countries to strengthen the medical education systems and increase the quality, quantity and retention of health care workers and the faculty needed to train them.

Information and Communication Technology (ICT)

Two more journals joined the African Medical Journal Editors Partnership Program (AJPP), developed in collaboration with the NIH Fogarty International Center in 2003, bringing the total to eight. The program is designed to strengthen African medical journals to enable their acceptance into Medline through capacity building and partnerships with six major medical journals in the US and UK. Through these development efforts, this program makes important research being carried out in endemic countries available to the world. NLM is the principal funder of the project.

Network of African Medical Librarians (NAML)

NLM continues its commitment to strengthening and a growing Network of African Medical Librarians (NAML) who received training as NLM Associate Fellows (http://www.nlm.nih.gov/about/training/associate/africannetwork.html) or as Medical Library Association Cunningham Fellows. The objective is to help African librarians, who already have links to NLM, build library capacity through outreach and training in Africa. The current network consists of eight librarians, from Kenya, Zambia, Mozambique, Mali, Nigeria, Morocco, Uganda, and Zimbabwe.

With the support of leaders of their medical schools, the African librarians have developed a course for the medical school curriculum of their respective universities as well as a training manual for use in Schools of Health Sciences in Africa called *Finding, Organizing, and Using Medical Information: A Training Manual for Students, Researchers and Health Workers in Africa.* Available in print and electronic formats, including YouTube, and 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 training course is also being used outside of the original institutions involved. OHIPD supports health information resource training provided by NAML librarians at meetings of the Association for Health Information and Libraries in Africa AHILA and at an annual Bioinformatics and Medical Informatics Workshop in Morocco.

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

Web Sites

NLM hosts a Malaria Research Resources Web site (http://www.nlm.nih.gov/mimcom/mimcomhomepage.html), which supports the activities of MIMCom, a project of the Multilateral Initiative on Malaria and the National Library of Medicine to support African scientists and malaria researchers. It also maintains a Web site that lists resources for International Librarians, Health Professionals and Researchers in Developing Countries.

Trans-NIH Collaboration

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

International Visitors

In FY2013, 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 62 countries:

Albania, Algeria, Argentina, Australia, Bahrain, Bangladesh, Bolivia, Brazil, Burma, Canada, China, Colombia, Costa Rica, Cuba, Czech Republic, Egypt, Fiji, Gambia, Germany, Guatemala, Haiti, Hungary, India, Indonesia, Iraq, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Korea, Kuwait, Laos, Lebanon, Liberia, Malawi, Mexico, Moldova, Nepal, Netherlands, Nigeria, Norway, Oman, Pakistan, Palestinian Territories, Peru, Qatar, Saudi Arabia, South Africa, Spain, Sri Lanka, Swaziland, Togo, Trinidad and Tobago, Tunisia, Turkey, Ukraine, United Arab Emirates, United Kingdom, Vietnam, Yemen.

Trans-NIH Collaborations in Data Science

The use of sophisticated computational approaches to explore and analyze vast amounts of complex, diverse data has transformed the modern world and is increasingly central to biomedical research. Key digital resources built and sustained by NLM, including PubMed/MEDLINE, GenBank, PubChem, and dbGaP, are critical to making important research results available and usable by researchers to fuel additional discoveries. In FY2013, the NIH launched a trans-NIH Big Data to Knowledge (BD2K) initiative designed to expand the capacity of the biomedical research enterprise to organize, sustain, analyze, integrate, and generate new discoveries from tidal waves of data emanating from research, health care, and social media. Currently, this initiative is planned as a seven-year endeavor., designed not only to fund data science research, infrastructure, and training, but also change NIH policies and practices to promote enhanced data management and access.

OHIPD played a leading role in overall direction of the Trans-NIH BD2K effort, as member of its Executive Committee. OHIPD and other NLM staff members also participated in several of the BD2K working groups and were instrumental in organizing several workshops of outside experts and in drafting requests for information and funding announcements.

The structural changes in the enterprise that BD2K hopes to achieve are expected to add value to the NIH research investment and accelerate the pace of progress.

LIBRARY OPERATIONS

Joyce E. B. Backus
Associate Director for Library Operations

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).

Most LO activities are 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 these systems.

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.

Program Planning and Management

Priorities for LO programs are based upon the goals and objectives identified in the NLM's Long Range Plan for 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 from scientific data to published literature readily available to patient and consumer health information. In FY2013, LO working groups continued 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 21st Century Workforce.

TSD proposed and drafted a division-wide reorganization to merge acquisitions and processing of serials and monographs and integrate all systems support functions. To support the reorganization and utilize space more efficiently, staff developed a plan for consolidation of all TSD division staff on a single floor of the NLM building.

LO continued its work toward developing a 21st century workforce by fostering professional growth and life-long learning of personnel.. In response to the results of a training needs survey, LO purchased a site-wide license for a commercial online training resource which provided access to over 2,000 video and hands-on courses. LO also continued to encourage its staff to take advantage of flexiplace work arrangements as appropriate. One hundred and forty-two LO employees work at home at least one day per week and another 68 have ad hoc telework agreements in place.

Collection Development and Management

NLM's comprehensive collection of biomedical literature is the foundation for many of the Library's services. LO ensures that this collection meets the needs of current and future users by updating NLM's literature selection policy; acquiring and processing relevant literature in all languages and formats; organizing and maintaining the collection to facilitate current use; and preserving it for subsequent generations. At the end of FY2013, the NLM collection contained 2,752,290 volumes and 19,056,616 other physical items, including manuscripts, microforms, prints, photographs, audiovisuals, and electronic media.

Selection

The Library's monographic and audiovisual collections were enriched by several notable donations, including

those from the American Medical Association, the National Institute on Aging's Alzheimer's Disease Education and Referral Center, the Navy's Bureau of Medicine and Surgery Office of the Medical Historian, the Armed Forces Medical Library, Pan American Health Organization, National Science Foundation Library, and the Public Health Service Historian's Office. Following review by NLM staff, 643 donated titles were added to the collection in FY13. Many of these works are out of print and are not widely held by other libraries. Staff worked on filling in gaps in the collection, for example, U.S. hospital, laboratory, medical school and medical society histories. These works are important to document the history of medicine, but often are difficult to acquire. As a result of this effort, the Library acquired approximately 50 retrospective titles for its collection in FY2013.

Acquisitions

LO received and processed 116,719 contemporary physical items (books, serial issues, audiovisuals, and electronic media). The total number of electronic-only serials grew to nearly 2,800 by the end of FY2013; electronic-only serials now represent more than 15 percent of all currently acquired serials. In FY2013, 6,381 licensed and 4,777 free electronic journals were available to NLM users. A net total of 23,628 volumes and 916,034 other items (including non-print media, manuscripts, and pictures) were added to the NLM collection.

Additions to NLM's world-renowned historical collection included:

- Four first edition prints from the Bernhard Siegfried Albinus monumental atlas, *Tabulae Sceleti et Musculorum Corporis Humani (Tables of the Skeleton and Muscles of the Human Body)* printed in Leiden between the years 1740 and 1742; these prints are the generous donation of Gloria and Paul Spiekermann of Westport, Connecticut;
- Guenther, Johann, von Andernach. *Institutionum Anatomicarum Secundum Galeni sententiam ad candidatos Medicine Libri Quatuor…ab Andrea Wesalio Bruxellensi, auctiores & emendatiores redditi.* [Venice: Melchiore Sessa, ca. 1538-1540]. This book is the second edition of Vesalius' extensively revised pocket edition of his teacher's anatomical handbook;
- Johannes de Ketham, *In disem biechlin find ma[n] gar ain schöne underwysung un[d] leer wie sich die Cyrurgici oder wundartz gegen ainen jeglichen verwundten menschen...* [Cologne: Aich, (1515?)]; this book is a very early edition of the translation into German of Ketham's *Fasciculus medicinae*, the first printed book to contain anatomic illustrations; and
- Weickard, Arnold. *Pharmacia Domestica, Dass ist: Hauss-Apoteck…* (Frankfurt am Main: E. Kempffer, 1628), an extensive collection of pharmaceutical recipes created for household preparation and use.

Unique additions to the archival and manuscript collections included the papers of:

- Bernadine Healy, from her tenure as Director of the American Red Cross, a gift of her husband, Floyd Loop;
- Larry Fagan and Ted Shortliffe, pioneers in the medical informatics field;
- Franklin Neva, renowned virologist, parasitologist, clinician and former chief of the Laboratory of Parasitic Diseases at the National Institute of Allergy and Infectious Diseases; Albert Kapikian, a virologist with the National Institute of Allergy and Infectious Diseases; and
- Tod Mikuriya, psychiatrist and grandfather of the medical Cannabis movement in the United States.

Unique additions to the historical audiovisuals collections included:

- 20 titles from the Pan American Health Organization, many focused on emergency preparedness and on AIDS;
- 28 films and several tapes documenting the work of Dr. Charles Hufnagel in developing a prosthetic heart valve and performing cardiac surgery, donated by his daughter Katherine Hufnagel; and
- 132 veterinary medicine titles from the University of Missouri, Zalk Veterinary Medical Library.

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. 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.

Collection and Space Management

In FY2013, LO bound 14,903 volumes, repaired 994 items in NLM's onsite book repair and conservation laboratory, made preservation copies of 632 motion pictures and videos, and completed conservation treatment of 375 Items specifically for digitization. Seventy-six US libraries donated 1,987 issues or volumes of serials to fill gaps in the NLM collection.

Preservation and Collection Management librarians worked with History of Medicine Division conservation staff to provide specialized salvage and recovery training for rare book, audiovisual, photographic and exhibition collections. Staff also participated in disaster response exercises as part of LO's Continuity of Operations Plan (COOP).

Work continued on the multiyear project to improve building infrastructure in the stacks and expand existing collection space. HMD's 19th century collection moved into new compact shelving on a floor recently strengthened to accommodate it. A new sprinkler system was installed overhead bringing this collection space to current fire and life safety codes. Additionally, HMD's oversize print collection housed in map cases was moved onto compact shelving rails. Collections stored temporarily at NLM's offsite location were returned to the Library. By the end of FY2013, all of NLM's collections were once again back in the building, with the exception of some unprocessed manuscripts housed at the University of Maryland.

National Cooperative Preservation

MedPrint, a national cooperative medical serials print retention program launched in September of 2011, has grown closer to meeting its goal of 13 copies of 250 key biomedical journals held in geographically dispersed locations throughout the country. The program, designed to ensure the preservation of and continued access to selected biomedical journals in print, is open to all US libraries in DOCLINE.

In all 68 libraries have committed to holding 772 titles, including all but six titles on the original list of 250. MedPrint participants voluntarily agree to retain titles for up to 25 years, or until 2036. The MedPrint Web site: http://www.nlm.nih.gov/psd/printretentionmain.html includes detailed information about the program, including links to the downloadable title list, reports of titles with commitments and the formal agreement which is to be signed by the participating institution and NLM.

Digitization Program

As part of its ongoing effort to make historical medical content more widely available to the public, in FY2013 NLM digitized 2,235 titles, consisting of 540,830 pages, including, among others, materials focused on World War I, in preparation for the 100th anniversary of the outbreak of the war, NLM's own publications, and materials from an NLM exhibition. Selected rare titles and manuscripts were scanned from the "Shakespeare and the Four Humors" exhibit. NLM's publications digitized included of the entire *Index-Catalogue of the Library of the Surgeon General's Office.*

In May, NLM began a "scan on demand" program. Patrons who request monograph material published before 1923 that is in the public domain will receive a fully digitized copy of the original, which is then also deposited in NLM's Digital Collections repository.

Permanent Access to Born Digital and Digitized Information

The preservation of electronic information presents unique challenges that are not yet fully understood. 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.

The NLM repository, Digital Collections, now holds almost 10,000 monographs and serials and over 100 films. A new release of Digital Collections included an upgrade to the Fedora Commons framework, keeping the repository's technology within active community support and development. A redesigned homepage with informative images highlighting repository content and responsive sizing of homepage and search results to accommodate the wide range of displays better was implemented. The Muradora front-end application was replaced with Blacklight, an open-source discovery interface which provides a new "Refine by" feature on the left allowing users to limit searches to specific facets.

NLM released to the public a Health and Medicine Blogs Collection and subsequently established an NLM program of Web collecting. Program staff are developing regular processes and workflows for selecting, acquiring and preserving born-digital Web content.

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 clinical vocabularies designated as US standards for use in patient records and health information exchange; and works with OCCS and LHNCBC to produce the Unified Medical Language System (UMLS) Metathesaurus, 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.

Staff in the National Information Center on Health Services Research and Health Care Technology (NICHSR), in MeSH, and in BSD contribute many interagency standards development and policy efforts.

In FY2013, NLM established a follow-on interagency agreement (IAA) with the Centers for

Medicare and Medicaid Services (CMS) to provide funds to NLM for (1) the establishment of infrastructure and tools for the validation, storage, dissemination, authoring, and maintenance of quality measure value sets used in Meaningful Use; (2) actual validation, engagement with measure developers to correct problems identified, and dissemination of standardized value sets for quality measures required for Stage 2 Meaningful Use and beyond; and (3) appropriate maintenance and additions to clinical terminology standards as necessary to standardize quality measure value sets. The FY2012 IAA and this follow-on IAA contribute substantially to the development by NLM of the Value Set Authority Center (VSAC).

In FY2013, NLM also established a follow-on agreement with the Office of the Assistant Secretary for Planning and Evaluation (ASPE) to support work in multi-year collaboration with the ASPE and the Office of the National Coordinator (ONC), intended to help bolster US data infrastructure for patient-centered outcomes research (PCOR) through identification, promotion and use of common data elements (CDEs) in clinical research. Initial NLM activities related to this agreement included the establishment of a working group under the Trans-NIH Biomedical Informatics Coordinating Committee, launch of an NIH CDE activities Web portal (www.cde.nih.gov), identification of standard grant language that would require consideration and use of relevant CDEs in NIH-funded research, and development of a prototype tool for storing and retrieving CDEs consistent with VSAC and other NLM terminology resources. NLM also participated in ONC's companion activity to identify standards and guidelines for use of standardized forms and data elements in electronic health records (EHR) systems to support structured data capture for uses other than clinical care, such as PCOR, adverse event reporting, and public health.

Also in FY2013, NLM established an interagency agreement with the US Department of Veterans Affairs (VA) to accelerate the pace of clinical terminology standards development and integration in areas that support Veterans health care and benefits determination.

Medical Subject Headings (MeSH)

The 2013 edition of MeSH contains 27,149 descriptors or main headings, 83 subheadings or qualifiers, and more than 219,266 supplementary concept records (SCRs) for chemicals, other substances, and diseases. For the 2013 edition, the MeSH Section added 304 new descriptors, replaced 40 descriptors with more up-to-date terminology, and deleted 8 descriptors.

This year MeSH began a Unique Ingredient Identifiers (UNIIs) Project. The project is a collaborative effort between NLM and the Food and Drug Administration (FDA) to include FDA Substance Registration System (SRS) - Unique Ingredient Identifiers (UNIIs) in MeSH. The first part of this project, completed for the 2013 MeSH year, involved updating MeSH SCRs with UNII identifiers. The 2014 MeSH release will complete the second part and include 2,900 UNII updates for Descriptor chemicals. In total, there will be 10,837 UNII identifiers in MeSH.

Because many of the MeSH UNII concepts involve pharmaceutical compounds, the range of pharmacological action (PA) classes covered by MeSH was expanded and a new category, "Diagnostic Uses of Chemicals," was added. The addition of UNII identifiers and new MeSH PA assignments has the potential to improve the utility of the MeSH database as a resource for drug information.

MeSH finalized the integration of Genetics Home Reference (GHR) disease vocabulary completing a goal of covering more than 95percent of all the human diseases and conditions with genetic components in MeSH. Human disease terms from the NIH Office of Rare Diseases and Research (ORDR), Online Mendelian Inheritance in Man (OMIM), and GHR are now part of MeSH.

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 FY2013, the Metathesaurus contained more than 2.9 million concepts and 11.4 million concept names from 168 source vocabularies in 19 different languages. The following new sources and updates were among the changes to the UMLS in FY2013: Ages & Stages Questionnaire; Pressure Ulcer Stages; Veterans RAND 12 Item Health Survey (VR-12); Source of Payment Typology (SOP5), Current Procedural Terminology (CPT2013); International Classification of Diseases, 10th Edition, Clinical Modification (ICD10CM2013); Logical Observation Identifiers Names and Codes (LOINC); Medical Dictionary for Regulatory Activities Terminology (MedDRA) 15.1 (plus translations); MEDCIN; Medical Subject Headings; RxNORM, and SNOMED CT.

Clinical Vocabularies

NLM is the central coordinating body for clinical terminology standards within the HHS. LO, in partnership with LHNCBC and OCCS, represents NLM in Federal initiatives to select and promote use of standard clinical vocabularies in EHRs 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 continue to intensify, particularly in the areas of quality and performance measurement, lab services, and newborn screening. MeSH vocabulary work is intimately tied to NLM's Value Set Authority Center (VSAC) work, which supports the Meaningful Use initiatives of CMS and ONC. VSAC, launched in October

2012, provides downloadable access to all official versions of vocabulary value sets contained in the 2014 Clinical Quality Measures. The value sets deliver lists of the numerical values and individual names from standard vocabularies used to define the clinical concepts (e.g., diabetes, clinical visit) used in the quality measures. The content of the VSAC will gradually expand to incorporate value sets for other use cases, as well as for new measures and updates to existing measures.

RxNORM

The MeSH Section produces RxNorm, a clinical drug vocabulary that provides standardized names for use in prescribing medications and a mechanism for connecting information from different commercial drug information sources. RxNorm is designated as a US clinical vocabulary standard by the Secretary of Health and Human Services. In FY2013, the Anatomic Therapeutic Class (ATC) drug classes published by the World Health Organization (WHO) were added to RxNorm. In addition, a number of large scale data fixes and improvements related to brand names, current on-the-market drugs and FDA ingredient codes were made to RxNorm data.

LOINC

NLM supports the continued development and free distribution of LOINC® (Logical Observation Identifiers Names and Codes) by the Regenstrief Institute at Indiana University. LOINC is one of the required standards for use in EHRs to achieve Meaningful Use, as defined by the ARRA-HITECH provisions. This year, an agreement was made to begin work to link LOINC and SNOMED Clinical Terms which will help improve safety, functionality and interoperability of health data within electronic medical records.

SNOMED CT

In FY2013, 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. SNOMED CT continues to be identified as one of the required standards for use in EHRs to achieve Meaningful Use, as defined by the ARRA-HITECH provisions. FY2013 saw several important activities that will improve the use of SNOMED CT in the US and elsewhere.

NLM continues as the lead in the development of mappings from standard vocabularies to administrative code sets (e.g., SNOMED CT to ICD-10-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. In FY2013, NLM continued to participate in several mapping projects, both nationally and internationally, to achieve these goals. Most importantly, NLM published a map from ICD-9-CM Procedure Codes to SNOMED CT to support the transition from use of legacy ICD-9-CM procedure codes to SNOMED CT. NLM also updated the ICD-9-CM Diagnostic Codes to SNOMED CT Map, as well as the SNOMED CT to ICD-10-CM Map.

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. Health science libraries world-wide use the cataloging and authority data created by NLM, thereby minimizing the cost and effort of cataloging their collections. Cataloging is performed by the TSD Cataloging Section, staff in HMD, and contractors.

The Cataloging Section is responsible for maintaining and updating of the *NLM Classification*, the maintenance and updating of the NLM Catalog Document Type Definition (DTD), transformation of Machine-Readable Cataloging (MARC) data into other metadata formats, and name authority control. The Cataloging Section released the 2013 edition of the *NLM Classification* in online and PDF formats providing updates to the QU (Biochemistry) schedule.

On March 31, 2013, NLM implemented the new cataloging guidelines, RDA (Resource Description and Access) for all original cataloging. Cataloging Section management developed and provided numerous training sessions on RDA for staff throughout the Library. NLM has shared with the community some of the local documentation used by NLM catalogers when creating RDA records. The Library of Congress (LC) and other national libraries also implemented RDA at the same time.

Using data provided by LC, NLM undertook a special project to add LC identifiers wherever possible to

names in the NLM authority files. This project allowed NLM to update over 400,000 name authority headings affected by RDA changes and the associated bibliographic records in an automated fashion.

NLM participated in the BIBFRAME Early Experimenters (EE) group assembled by the Library of Congress to replace the MARC 21 format. BIBFRAME, or the Bibliographic Framework Initiative, is a community undertaking to determine a transition path for the MARC 21 exchange format to more Web-based, Linked Data standards.

In FY2013, the Cataloging Section cataloged a total of 16,685 books, serials, electronic resources, and audiovisuals. These figures include 1,959 Cataloging-in-Publication (CIP) records created from publisher electronic galleys or ONIX data and 1,504 titles cataloged for the historical collection. The FY2013 figures are lower than the previous year due to time needed by staff to learn and understand RDA, as well as a significant decrease in the number of CIP titles being received from the Library of Congress.

Indexing

LO indexed 5,640 biomedical journals for the MEDLINE/PubMed database to assist users in identifying articles on specific biomedical topics. In FY2013, a combination of in-house staff, contractors, and cooperating US and international institutions indexed 734,052 articles, a 4 percent decrease from last year. Previously indexed citations were updated to reflect 467 retracted articles, 9,178 published corrections, and 48,714 comments found in subsequently published notices or articles.

In FY2013, indexers created 85,429 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI Gene database. By the end of FY2013, 4,273 MEDLINE journals (76 percent) were indexed from the online versions, accounting for 88 percent of all new citations.

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 LHNCBC and OCCS on the Indexing 2015 initiative to identify, test, and implement ways to reduce or eliminate tasks now performed by human indexers. In FY2013, continued progress was made with the Medical Text Indexer First Line (MTIFL) 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 FY2013, 7,956 articles from 88 journals had been processed using MTIFL for the year.

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 FY2013, 94 percent of the citations and abstracts were received electronically from publishers and the remaining 6 percent were scanned.

The Index Section expanded automated comment indexing to cover comment articles regardless of where the referent article is published; previously comment indexing and links were made only between articles and comment articles within the same journal. For FY2013, the Index Section applied automated indexing to 31,018 comments, an increase of nearly 4 percent. In addition, author replies to comments citations are now cited as separate citations, and indexed as a regular type of comment article.

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 FY2013, the Committee reviewed 517 journals. They recommended a total of 67 journals, including 23 titles that were recommended provisionally as they are available only in electronic form and must meet the MEDLINE electronic journal policy before indexing begins.

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 the MEDLINE/PubMed NLM's database of indexed journal 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. These databases are richly interlinked with each other and with other important NLM resources, including PubMed Central, 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 2.5 billion searches, a 14 percent increase from last year. Page views in PubMed totaled 7.57 billion. MEDLINE/PubMed now includes more than 23 million citations. For FY2013, NLM allowed publishers providing citation data to submit several new types of data. They may now submit author-supplied keywords, as well as the article publication type Review and non-English abstracts as secondary abstract content available on the publisher's Web site.

DailyMed® provides access to over 58,000 structured product labels (SPL) from the Food and Drug Administration. The ability to browse drug classes was added to the DailyMed Web site, and DailyMed had more than 146 million page views in FY2013, nearly double from the previous year.

LO 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 FY2013. Over 298 million visitors viewed over 845 million pages on the MedlinePlus full Web site. MedlinePlus Mobile received over one million visitors viewing over 4.5 million pages. At the end of FY2013, there were 660,000 subscribers with over 65.2 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.

MedlinePlus Connect, a free service that brings patients or providers in an EHR system or patient portal to targeted consumer health information from MedlinePlus, continued to add users, both health care organizations and electronic health record systems. The service recorded over 5.5 million page views in FY2013. The MedlinePlus Twitter account, @medlineplus, reached over 64,000 followers at the end of FY2013, while the Spanish @medlineplusesp reached nearly 14,000. MedlinePlus grew by 22 English health topics and 21 Spanish health topics offering a total of 953 English health topics and 939 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 one topic in FY2013 bringing the total to 52 topics. The new topic was Healthy Eyes from the National Eye Institute.

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.

Full-text digitized resources on Bookshelf for health services research and public health continued to expand in FY2013. Now managed by NCBI, Health Services and Technology Assessment Text (HSTAT) on the Bookshelf added 98 new documents from the Agency for Healthcare Research and Quality (AHRQ), from its Effectiveness Program and Evidence Based Practice Centers (Evidence Reports/Technology Assessments series). Also new to HSTAT in FY2013 were *Health United States, 2012* from the National Center for Health Statistics: *Behavioral Health, United States, 2012* from the Substance Abuse and Mental Health Services Administration; and three reports on Hepatitis C, PTSD, and Caregiver Interventions from the US Department of Veterans Affairs. Overall, over 150 titles on comparative effectiveness research (CER), evidence-based medicine (EBM), health policy, HSR and public health (PH) from a variety of sources were added in to Bookshelf FY2013.

In FY2013, the HSR Info Central (HSRIC) site added three new topic pages, representing identification of current valuable grey literature resources: Evaluation Resources for Assessing HIT Systems and HIT Implementation Adoption and Use; Health Literacy and Cultural Competence; and Privacy/Security and Research with Electronic Health Records. As a result of these efforts, as well as increased promotion of the site to the HSR community, visits to HSRIC increased over 50 percent and visitors increased more than 35 percent in FY2013.

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 the HSRR (Health Services and Sciences Research Resources) database to include PHSSR (Public Health Systems and Services Research) datasets. In FY2013, the HSRProj database grew to almost 11,000 active records representing 149 funders, both national and international. In FY2013, the HSRR database increased to over 1,276 records describing datasets, survey instruments or software, of which almost 20 percent are explicitly related to public health systems and services research.

NICHSR continued to be involved in a number of activities related to examining how the NLM resources could support increasing interest in CER and Patient Centered Outcomes Research (PCOR). In FY2013, NICHSR and ADLO staff continued to serve on behalf of NLM as key technical advisors to ASPE on resources for CER. In FY2013, visits to NICHSR-developed and tested Comparative Effectiveness Research search strategies increased more than 9 percent.

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 767 MEDLINE licensees in FY2013, an increase of 4 percent from the previous year. Of the 808 total licensees of MEDLINE and other NLM databases, 582 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 FY2013, there were 8,804 UMLS Metathesaurus licensees, an increase of 56 percent 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 interlibrary loan services available from other libraries. In FY2013, PSD's Collection Access Section processed 294,483 requests for contemporary documents.

The number of onsite requests in the NLM Main Reading Room for items from the closed stacks continued to decline primarily due to online access to more journals in the Reading Room, and fewer onsite visitors. Onsite users requested 76,215 contemporary items from the stacks, an 11 percent decrease from last year. Reading Room staff continued to focus on providing optimal service to onsite users while streamlining and reducing unnecessary equipment and furniture. Users of the HMD Reading Room requested 7,639 items from historical and special collections.

In FY2013, the Collection Access Section processed a total of 218,268 interlibrary loan requests (ILL), a 7 percent decline from the previous year, continuing the decade long trend of decreasing demand for ILL service. 99 percent of all requests were filled within one day and NLM delivered almost all, 97 percent, of ILLs electronically.

A total of 2,558 libraries use DOCLINE, NLM's interlibrary loan request routing and referral system. NLM released six minor upgrades to DOCLINE during FY2013, including needed Section 508 compliance, changes to accommodate PubMed XML 2.0 data structure, and improvements to reports.

DOCLINE users entered 1.4 million requests in FY2013, a 9 percent decrease from the previous year; 93 percent of the requests were filled. The NLM share of all DOCLINE requests comprised 15 percent of the total. Individuals submitted 249,128 requests to DOCLINE libraries via the Loansome Doc feature in MEDLINE/PubMed, a 9 percent decline from the previous year. Document request traffic continues to decline in all Regions of the NN/LM due to expanded availability of electronic journals.

DOCLINE requests are routed to libraries automatically based on holdings data in its serial holdings database. At the end of FY2013, the holdings database contained over 1.6 million holdings statements for 62,247 serial titles held by 2,558 libraries. Electronic holdings comprise 33 percent of the total holdings.

In FY2013, NLM continued to examine the decline of DOCLINE use and researched the needs of resource sharing librarians in the national network. NLM surveyed 2,400 librarians in March to learn more about these issues. The survey achieved a 72 percent response rate, with results released at the annual Medical Library Association (MLA) meeting in May. Generally and as expected, the survey revealed that interlibrary loan is being supplanted by access to electronic journals.

NCBI, LO and staff at the Regional Medical Libraries (RML) 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. For FY2013, 2,987 libraries participate in LinkOut, a 7 percent increase since last year. There were 1,085 libraries participating in the Outside Tool option, a 13 percent increase.

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 require 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 107,965 inquiries from the public and libraries in FY2013 and answered 5,219 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 called 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

6,177 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 FY2013, the Center focused on supporting the Regional Medical Libraries (RMLs) in planning and evaluating four major outreach initiatives: outreach to community colleges; outreach to K-12 communities; ClinicalTrials.gov Results outreach; and outreach to support MedlinePlus Connect adoption. 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. During 2013, the focus was 1) continuously monitoring and improving performance of the nnlm.gov environment, 2) assisting NN/LM staff to achieve improved compliance with Section 508, and 3) building Drupal architecture for the future iteration of nnlm.gov. 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. In all, the NN/LM funded 206 outreach projects in FY2013. 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, 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 FY2013, NN/LM services were exhibited at 390 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 FY2013, the Web site expanded significantly with more than 946 new links.

One of the most popular resources on the PHPartners site continues to be the pre-formulated PubMed search queries in support of the HHS Healthy People 2020, which launched in 2010. In FY2013, NICHSR continued to work with the HHS Office of Disease Prevention and Health Promotion to refine the structured evidence queries or "SEQs," working with librarians and public health subject matter experts to develop additional queries for each of the Healthy People 2020 objectives and sub-objectives. By the end of the FY2013, SEQs for 30 HP2020 topics had been posted, with work underway for release of the remaining SEQs by mid FY2014. The SEQs are accessible from both the PHPartners Web site and from the Healthy People 2020 Web site, at www.healthypeople.gov. The use of SEQs from the PHPartners Web site increased by 28 percent in FY2013.

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 several years, LO has worked in collaboration with NLM's Director of International Programs to improve health information capacity in sub-Saharan Africa.

LO worked with OHIPD to continue 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.

Historical Exhibitions and Programs

The History of Medicine Division's Exhibition Program directs the development, installation, and travel of historical exhibitions designed to appeal to the interested public as well as the specialist. These exhibitions highlight the Library's historical resources, foster partnerships with other federal agencies and cultural organizations, and are an important part of NLM's outreach program. Complementing these exhibitions are a number of programs of the History of Medicine Division designed to raise scholarly, educational, and public awareness and use of the NLM's rich historical collections which date from the 11th century to recent decades.

The Exhibition Program provided tours to 1,982 visitors to the NLM's *Native Voices: Native Peoples' Concepts of Health and Illness* exhibition. Staff continued to support the Library-wide effort to develop and travel a banner/iPad adaptation of the exhibition to four sites. The Exhibition Program traveling exhibition services continued to expand. In FY2013, 25 banner exhibitions traveled to 40 states, one US territory, and three international sites for a total of 145 exhibition openings. All shows are booked into 2016 with some itineraries extending into 2018. Based on preliminary returns from a soon to be completed survey, over 400,000 people have seen a National Library of Medicine exhibition in FY2013.

In partnership with the National Museum of American History, the Exhibition Program developed *From DNA to Beer: Harnessing Nature in Medicine and Industry*, a special display, traveling banner exhibition, and online exhibition with education resources about the history of scientists and industry using microbes to develop new therapeutics. The project showcases items from the History of Medicine Division collections and the National Museum of American History.

Surviving and Thriving: AIDS, Politics, and Culture, a traveling banner exhibition and online project, which the Exhibition Program produced, includes over 250 public health posters about HIV/AIDS drawn from the Images and Archives collections. The education resources for this project include two lesson plans, three higher education modules, and a new feature: Related Resources at NLM.

In collaboration with colleagues from the Digital Manuscript Program, the Exhibition Program produced three lesson plans and one higher education module in support of the Profiles in Science site: The Charles R. Drew Papers.

Noteworthy cooperation between the History of Medicine Division and external partners, included:

- Two loans of books from the NLM collections to The Cloisters, a branch of the Metropolitan Museum of Art in New York, for the exhibition, "Search for the Unicorn: An Exhibition in Honor of the Cloisters' 75th Anniversary," and to The Grolier Club of New York for the exhibition, "Extraordinary Women in Science and Medicine: Four Centuries of Achievement." NLM staff worked with these institutions to disseminate news releases about these collaborations, and the blog *Circulating Now* also featured this content.
- The digital release of "The Reward of Courage," a 1921 silent film long believed to have been lost, in cooperation with the Library of Congress. This release coincided with the launch of *Medical Movies on the Web*, a curated portal featuring selected motion pictures from the NLM's world-renowned collection of over 30,000 audiovisual titles.
- Releasing the Web version and iPad App of the NLM's Turning the Pages initiative, featuring Elizabeth Blackwell's *A Curious Herbal*, in cooperation with the LHNCBC.
- Curating a new Profiles in Science site on Sir William Osler, who is recognized as one of the founders of modern science-based medicine, in collaboration with the Alan Mason Chesney Medical Archives of the Johns Hopkins Medical Institutions and the Osler Library of the History of Medicine at McGill University. Staff developed a Pinterest page for Profiles in Science as a means to share the images and stories of the history of medicine with new audiences. FY2013 marked the fourteenth year of Profiles in Science. To date, the program has digitized more than 190,000 pages of material.

This year marked a period of substantial change in the NLM's History of Medicine Division, with its staff collaborating in unprecedented ways across the Library to serve an expanding audience of scholars, educators, students, and members of the general public. Leading this change was the launch of the blog, *Circulating Now* (http://circulatingnow.nlm.nih.gov/), in cooperation with NLM's Office of Office of Communications & Public Liaison. Designed for a general audience, *Circulating Now* offers a variety of rich content about the NLM's historical collections and the opportunity to submit comments and exchange ideas about the subjects and themes of the posts. Kicking off *Circulating Now* was a series of posts that reenacted the medical account of the shooting and subsequent illness and death of President James A. Garfield during the summer of 1881. The entire set of public bulletins about the president's health were republished, in real historical time, 132 years to the day and hour of those fateful events. The result was a dramatic achievement that drew attention to the history of medicine in a way that connected a 21st century audience to the suffering of America's 20th president and the larger story that unfolded in the days following his shooting by Charles Guiteau.

Also instrumental in the HMD's expanding reach was its collaboration with the National Endowment for the Humanities to support initiatives that bring together scholars, scientists, librarians, doctors and cultural heritage professionals from the humanities and biomedical communities. Based on this collaboration, the division supported the public, interdisciplinary symposium "Shared Horizons: Data, Biomedicine, and the Digital Humanities,"

which explored research methodologies and interpretations of evidence arising in this burgeoning area of biomedical-driven humanities scholarship. The NLM's primary contribution to this symposium was its announcement of the availability of the eXtensible Markup Language (XML) data from the NLM's IndexCat™ database, which includes the digitized content of more than 3.7 million bibliographic items from the printed, 61-volume *Index-Catalogue of the Library of the Surgeon-General's Office*, originally published from 1880 to 1961. The XML describes items spanning five centuries, including millions of journal and newspaper articles, obituaries, and letters, hundreds of thousands of monographs and dissertations, and thousands of portraits.

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 Library of Medicine Training Center (NTC) at 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.

In FY2013, the MEDLARS Management Section (MMS) and the NTC trained 1,470 students in 51 classes including PubMed and TOXNET. An average of 13,648 unique visitors used the PubMed Tutorial for approximately 149,564 page views each month.

LO continues to create and enhance distance education resources. There are now 58 Quick Tours covering products such as PubMed, NLM Catalog, ClinicalTrials.gov, LinkOut for Libraries, and the UMLS. NLM and the NN/LM are increasing the number of Webcasts to provide current information on NLM products and services as a cost effective and timely way to reach larger numbers of users. Distance education courses expanded to include PubMed for Trainers, and PubMed for Librarians (five distinct modules), Fundamentals of Bioinformatics, and Searching, A Librarian's Guide to NCBI. NTC, MMS and SIS worked together to develop a new distance education course version of the TOXNET training course. NICHSR continues to make available its suite of courses on health services research, public health, and health policy. LO posted 10 training videos on the NLM YouTube channel, covering topics such as PubMed, UMLS and LinkOut for Libraries.

The NLM Associate Fellowship program had seven participants in FY2013, with four first-year Fellows on site at NLM and three second-year Fellows at other libraries. LO placed these second-year fellows at the University of Utah; Inova Fairfax Hospital in Fairfax, VA; and the NIH Library. In September 2013, a new group of five first-year Fellows began their year at NLM and four 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 2013, NLM participated 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 two 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

Physical	***Total***[1]	***FY2013***	***FY2012***	***FY2011***
Monographs[2]				
Before 1500	597	0	0	1
1501-1600	6,056	3	0	7
1601-1700	10,346	3	0	5
1701-1800	272,738	3	4	21
1801-1870	256,717	27	37	66
1871-Present	872,474	11,292	13,287	14,650
Bound Serial Volumes[3]	1,464,490	12,650	17,025	21,924
Microforms[4]	606,084	8	10	15
Audiovisuals and Computer Software	92,961	1,701	1,310	1,224
Prints and Photographs	69,952	300	336	20
Manuscripts[5]	18,288,707	914,025	1,716,225	1,143,275
Withdrawn Items	(132,212)	(350)	(294)	(1,272)
Total items	21,808,906	939,662	1,747,940	1,179,936

Digital	***Total***[1]	**FY2013**	***FY2012***	***FY2011***
PubMed Central Articles	2,859,268	318,316	263,203	223,404
PubMed Central Titles[6]	1,393	244	252	203
Bookshelf Titles[7]	1,719	323	323	295
Digital Collections Repository[8]				
Texts[9]	9,559	2,580	5,075	1,406
Audiovisuals[10]	107	37	41	18

[1] Total: Numbers are cumulative as of the end of the fiscal year.
[2] Monographs: A bibliographic resource complete in one part or finite number of separate parts. Includes Americana, theses and pamphlets. Starting in FY2011 numbers for these materials are reported under monographs by publication year.
[3] Bound serial volumes: A serial is a continuing resource issued in separate parts with no predetermined conclusion. Bound serial volumes include serials bound, serials pamphlet bound and bound serial gifts.
[4] Microforms: Reduced size reproductions of monographs and serials including microfilm and microfiche.
[5] Manuscripts: Total manuscripts equivalent to 8,948 linear feet of material, multiplied by a common factor to provide an item number estimate.
[6] PMC Titles: Only fully deposited titles.
[7] Bookshelf Titles: Titles of books, reports, databases, documentation, and collections.
[8] Digital Collections Repository: Digitized content in the public domain. In the future will contain born digital items as well as reformatted items.
[9] Includes monographs and serials such as annual reports. Referred to as “Print Materials” on Digital Collections Web site.
[10] Referred to as “Films and Videos” on Digital Collections Web site.

Table 2: Collection Activities

Acquisitions and Processing	***FY2013***	***FY2012***	***FY2011***
Active Serial Subscriptions	18,343	19,184	19,731
Items Processed[11]			
Serial Pieces	99,891	101,294	117,091
Monographs (pre-1914)[12]	336	715	886
Monographs (1914-)	16,530	17,012	20,639
Audiovisuals[13]	683	1,335	N/A
Prints and Photographs	1,397	47,982	897
Total	137,180	168,338	159,244
Archival Materials Acquired			
Modern Manuscripts (in linear feet)	120	497	866
Expenditures			
Publications	$11,033,522	$10,207,330	$10,216,505
Rare Books, Manuscripts, and other Historical Materials	$299,948	$299,584	$299,794
Total[14]	$11,333,470	$10,506,914	$10,516,299
Preservation			
Volumes Bound	14,903	15,000	18,997
Volumes Repaired Onsite[15]	994	2,346	2,304
Audiovisuals Preserved	632	534	594
Historical Volumes Conserved	375	997	340
Pages Digitized[16]	540,830	643,372	787,865

[11] Items processed: Serial issues, monographs and nonprint receipts processed.
[12] Monographs (pre-1914) includes historical manuscripts (those written prior to the year 1600).
[13] Audiovisuals became a separate tracking category in FY2012. For prior year reports, Audiovisuals were grouped with Monographs (1914-).
[14] Used to be reported in "Publications" prior to FY2012 and "Rare Books" was a portion of the amount.
[15] Volumes repaired onsite: General Collection monographs and serials only
[16] Number excludes digitization projects not associated with the Digital Collections Repository, e.g., Profiles in Science.

Table 3: Cataloging and Indexing

Cataloging	***FY2013***	***FY2012***	***FY2011***
General Collection Items[17]	16,685	19,656	21,268
Historical Monographs (pre-1914)	4,342	4,238	1,128
Modern Manuscripts (in linear feet)[18]	303	1,050	843
Prints and Photographs[19]	2,083	2,281	704*
Historical Audiovisuals	894	992	105

Indexing	***FY2013***	***FY2012***	***FY2011***
Citations Indexed for MEDLINE	734,052	760,903	724,831
Journals Indexed for MEDLINE	5,640	5,633	5,559

*This figure corrected from previous reports.

[17] Items: Includes monographs, serials, nonprint and integrating resources.
[18] Number reflects manuscripts that are fully processed and have a catalog record.
[19] Number includes accessioned prints and photographs that are described by finding aids.

Table 4: Services to the Public

Document Delivery	***FY2013***	***FY2012***	***FY2011***
Interlibrary Loan Requests Received	218,268	234,662	244,701
Interlibrary Loan Requests Filled	179,941	194,255	200,581
General Reading Room Requests			
Received	76,215	85,352	119,209
General Reading Room Requests Filled	68,486	74,993	102,054
History of Medicine Reading Room			
Requests Filled	7,639	6,989	5,346*

Customer Service Inquiries	***FY2013***	***FY2012***	***FY2011***
Offsite Inquiries[20]			
General	99,857	100,023	97,796
History of Medicine	8,108	10,857	8,667
Onsite Inquiries[21]			
General	5,219	6,124	6,531
History of Medicine	12,183	14,763	15,057

Data Licensees	***FY2013***	***FY2012***	***FY2011***
MEDLINE	767	736*	688
UMLS	8,804	5,640	6,566

Tours and Visitors	***FY2013***	***FY2012***	***FY2011***
Exhibitions			
Visitors	2,662	2,769	2,522
Daily Tours			
Tours	111	122	100
Visitors	734	668	548
Special Tours			
Tours	88	63	71
Visitors	1,660	1,047	1,392

*These figures corrected from previous reports.

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

Table 5: Selected Web Resources

Resource	*FY2013*	*FY2012*	*FY2011*
ClinicalTrials.gov			
Number of Trials	157,013	133,291	118,682
Page Views[22]	1,145,603,153	1,474,668,417	1,347,564,098
Visitors[23]	11,058,828	11,201,854	9,518,503
DailyMed			
Number of Labels	55,190	39,232	31,000
Page Views	146,216,340	85,025,368	118,163,000
Visitors	17,541,406	8,812,640	5,166,000
Genetics Home Reference			
Summaries	2,125	1,907	1,622
Page Views	161,918,165	145,647,998	142,967,488
Visitors	9,719,450	6,382,520	4,819,287
Household Products Database			
Number of Products	12,000	11,000	11,000
Page Views	39,960,413	17,994,336	10,254,423
Visitors	820,718	786,346	694,042
MEDLINE/PubMed			
PubMed Citations	23,187,946	22,174,097	21,200,952
Page Views	7,573,459,637	6,113,756,836	4,783,804,040
Unique Visits[24]	716,895,000	482,175,373	539,184,717
Searches[25]	2,514,611,135	2,204,710,177	1,834,537,403
MedlinePlus			
Number of Topics (English/Spanish)	953/939	931/918	903/881
Page Views	845,000,000	745,000,000	728,000,000
Visitors	298,500,000	213,000,000	164,000,000
NLM Main Web Site			
Page Views	44,600,000	44,700,000	39,300,000
Visitors	9,100,000	8,900,000	8,300,000
ToxTown			
Page Views	5,029,713	6,757,011	6,164,031
Visitors	263,781	240,277	192,737

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

SPECIALIZED INFORMATION SERVICES

Steven Phillips, MD
Associate Director

The Division of Specialized Information Services (SIS) at the National Library of Medicine (NLM) develops and provides access to a diverse collection of free online information resources and services in toxicology, environmental health, chemistry, HIV/AIDS, health-related disaster management, minority health, and other specialized topics. SIS also develops and implements outreach activities to improve access to and use of these and other NLM resources. The SIS Outreach and Special Populations Branch actively works with underserved and other specific populations. SIS also manages NLM's Disaster Information Management Research Center (DIMRC).

We use many avenues of communication to reach our diverse user communities. The LactMed graphic above is one example of the way we have been able to use images on NLM's main Web site's homepage to highlight items and bring them to the attention of other NLM users. Many SIS toxicology and environmental health resources have appeared in this section, with high rates of interest from users as demonstrated by the "clicks" they received. We continue to provide Web resources in the area of toxicology and environmental health information, using social media methods that allow us to rapidly deploy new important relevant information on our own resources or from other federal partner agencies. We use a toxicology and environmental health listserv, Facebook, Twitter, 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.

SIS has played a leadership role 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 both educators and students with its K-12 initiative. The goal of these outreach efforts is to increase awareness and use of NLM's resources as well as to help users develop skills in critical assessment of health information resources.

This year DIRMC continued to expand its activities as well as respond to national and international crises, such as the Hurricane Sandy. The Chemical Hazards Emergency Medical Management (CHEMM) resource, a companion to REMM, released in FY2011, is now also included in WISER, the Division's versatile tool for assisting emergency responders. CHEMM was created to assist emergency managers, planners, responders and other healthcare professionals in responding to mass casualty chemical events.

During February-May 2013, the SIS Office of Clinical Toxicology (OCT) completed its fifth year of co-directing the Public Health Informatics course at the Uniformed Services University of the Health Sciences (USU). This course is jointly sponsored by the USUHS Biomedical Informatics Department and the National Library of Medicine and is intended 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. OCT staff completed a second year of lectures in the toxicology courses offered by the Preventive Medicine & Biometrics Department at USU. The topics included NLM's toxicology databases and other resources, toxicokinetics, risk assessment, risk management, snake and other venoms, and plant toxins.

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

Based on a user survey and the changing technical and budgetary environment, SIS decided to cease

maintaining the DIRLINE database. DIRLINE will remain searchable through FY2014 with the existing records marked as "archived." These changes will not affect Health Hotlines, which will be kept current.

Toxicology and Environmental Health Resources

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established more than 45 years ago within the National Library of Medicine as a result of recommendations of the President's Science Advisory Committee. From its inception, TEHIP has strived to use the most up-to-date 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, accurate 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 Laboratory Safety, Nanotechnology, and many resources related to health aspects of disaster preparedness and response.

TOXNET (TOXicology Data NETwork) http://toxnet.nlm.nih.gov is the core of the TEHIP program. The TOXNET databases are varied in what they cover, and how they are created and maintained. Some are legacies that are static while other have evolved from their earlier forms as we included information on a wider range of toxic chemicals, such as adding Superfund-cited compounds, focusing on worker exposures, taking on the possible safety issues with nanomaterials, considering household products safety. Others are databases originally created elsewhere and put up on the TOXNET system for greater visibility and to provide continued access when the project that created them was stopped. Enhancements to TOXNET continue, based on user feedback and upgrades or additions of data and capabilities. A new interface for TOXNET was in testing in FY2013, and is to be released in early 2014. A full list of TOXNET databases is available at http://toxnet.nlm.nih.gov/. They include:

LactMed (Drugs and Lactation), which provides information on drugs and other chemicals to which breastfeeding mothers may be exposed. It includes information on the levels of such substances in breast milk and infant blood, and the possible adverse effects in the nursing infant and includes links to other NLM databases. Complementary and alternatives medicines (CAM) are now included in LactMed, with a focus on supplements that are frequently used by nursing mothers. A LactMed app for the iPhone, iPod, iPad, and Android released last year continues to be very popular. LactMed has received several endorsements including one from the American Academy of Pediatrics: "The most comprehensive, up-to-date source of information regarding the safety of maternal medications when the mother is breastfeeding is LactMed," and from an FDA official in MedPage Today, August 26, 2013: "Pediatrician Hari Cheryl Sachs, MD, of the Food and Drug Administration, and colleagues suggested that healthcare professionals 'look up safety of specific medications on the National Library of Medicine's peer-reviewed LactMed database online.'

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 and includes structures from ChemIDplus displayed as images at the start of the record. Records continue to be added on EPA's green chemicals and pesticides, plant-derived toxins, and animal venoms. A related effort that began in 2013 was to look at substances "in the news" to identify substances that might be added to HSDB. Following the suggestions in the recent HSDB Needs Assessment project for the re-engineering of the HSDB, we continued to work to better incorporate new resources important to today's toxicologists as well as trying to identify significant new chemicals that should be included. We also began to transition the peer review process to one allowing more electronic reviews.

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 records on toxicology research projects from NIH's new system RePorter, allowing users to link directly back to the NIH database for full information on the projects.

Toxic Release Inventory (TRI), a series of databases that describe the releases of toxic chemicals into the environment annually for the 1987-2011 reporting years. TRI data is also available through TOXMAP (see below). It was updated with the most current year of data (2011), which is also reflected in TOXMAP.

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 387,000 chemical records, of which over 300,000 include chemical structures. ChemIDplus includes some toxicity data as well as locators to many important national and international listings of chemicals. ChemIDplus also serves as a locator to the chemicals in the other TOXNET databases. ChemIDplus serves as an essential authority component

for Toxnet and for the Drug Portal as well as providing data to PubChem.

Household Products Database, which provides information on the potential health effects of chemicals contained in more than 12,000 common household products used inside and around the home. New types of products were added this year. Collaboration continues via an Interagency Agreement with the Consumer Product Safety Commission (CPSC) to include various types of nanomaterials and additional categories of products in 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. This year we added 1,699 chemicals from the Department of Labor Site Exposure Matrices (SEM) database. In preparation for the April 2013 Isocyanates and Health: Past, Present, and Future international conference, the isocyanate-related content in Haz-Map was reviewed and updated. We continue to work with Department of Labor on an Interagency Agreement to include more frequent updates and add more chemicals from the SEM database. We also added three new fields and each chemical in Haz-Map was checked against the cancer designation of three organizations, International Agency for Research on Cancer (IARC), National Toxicology Program (NTP), and the American Conference of Governmental Industrial Hygienists (ACGIH).

ALTBIB is a Web portal providing information on resources about alternatives to the use of live vertebrates in biomedical research and testing, developed as part of NLM's participation in the Interagency Coordinating Committee on the Validation of Alternate Methods (ICCVAM). The new version of ALTBIB, released in September 2012, has been well-received within ICCVAM and beyond. ALTBIB provides easy access to key organizations, the latest news from some of those organizations, to U.S. and international compilations of validated methods and methods undergoing validation, and to numerous pre-formed searches of PubMed on alternative methods and other topics relevant to alternatives to the use of live vertebrates. Efforts are continuing to include pre-formed searches of the latest methods and topics of interest to ALTBIB's users.

In addition to the core TOXNET databases, SIS supports many other 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 FY2013 include updates to EPA Toxics Releases (2011 data) and EPA Superfund NPL data. Work was continued on a new version of TOXMAP which will more closely align with the look and feel of current mapping tools.

Drug Information Portal, providing current drug information for over 28,000 drugs with links to many credible additional online resources. During FY2013, changes were made to the search interface to allow easier selection of drug names and searching by category of drug.

Pillbox, is a new drug information resource that focuses on pill images as well as drug names and other physical appearance information. We worked 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. This year, we incorporated high-resolution images produced by NLM HPCC's Computational Photography Project for Pill Identification project and images from the Department of Veterans Affairs Medication Image Library. The collection has increased to more than 5,000 images of solid dosage pharmaceuticals. Additionally, the identification and search functionalities have continued to improve.

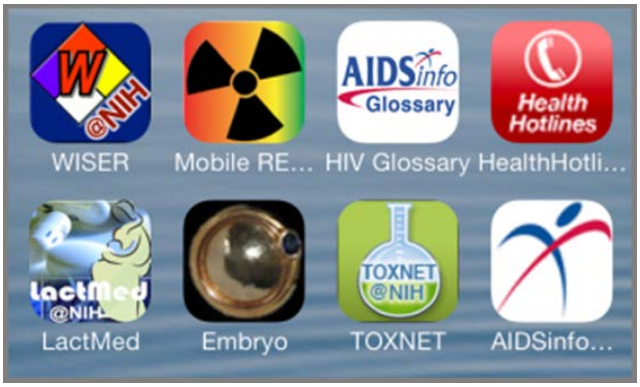
A new **Dietary Supplements Label Database**, created in collaboration with the NIH Office of Dietary Supplements (ODS), was released in June. This comprehensive labels database is intended primarily for researchers and will eventually include most of the estimated 55,000 dietary supplement products available to US consumers.

LIVERTOX provides comprehensive, evidence-based information on hepatotoxicity attributable to prescription and nonprescription medications, herbals and dietary supplements, and it includes an international clinical case registry. The LIVERTOX Web site is a resource for physicians, patients, clinical academicians and researchers that has become a high-value Web resource with about 150,000 visits per month, with the number rapidly trending upward. LIVERTOX content is produced by the NIDDK in collaboration with SIS/NLM, with approximately 750 drugs and herbals currently in the database. For FY2013, more than 150 drug records were added to the database, and more than 50 percent of the drug records were reviewed or enhanced with the latest references.

Mobile Resources and Apps

In response to the President's Digital Strategy, SIS explored and promoted the use of emerging mobile technologies to deliver health information from the new or existing resources in the hands of our users. In 2013, SIS enhanced 11 mobile apps/mobile Web resources in the

area of emergency response, health information, drug information, toxicology, and embryology.



WISER/CHEMM

WISER (Wireless Information System for Emergency Responders) is a mobile app designed to assist first responders in hazardous material incidents. In 2013, WISER integrated Chemical Hazards Emergency Medical Management (CHEMM) content. The app is available on iPhone, iPad, Android, and Blackberry.

REMM (Radiation Emergency Medical Management) is a mobile app designed to provide guidance for healthcare providers about clinical diagnosis and treatment of radiation injury during radiological and nuclear emergencies. REMM is available on iPhone, iPad, Android, and Blackberry.

Disaster Apps and Mobile Optimized Web Pages is a mobile Web resource that aggregates emergency response mobile tools. This resource was first released in 2013.

HIV/AIDS Glossary

The AIDS*info* HIV/AIDS Glossary is a tool that provides mobile access to over 700 pertinent HIV/AIDS glossary terms in both English and Spanish. It is available on iPhone, iPad, and Android.

AIDS*info* Mobile

This mobile site provides the latest federally approved HIV/AIDS medical practice guidelines, HIV treatment and prevention clinical trials, and other research information.

Health Hotlines

The Health Hotlines is a mobile app of health organizations with toll-free telephone numbers. The information is in English and Spanish. It is available on iPhone and iPad.

HazMap

The HazMap mobile Web site provides occupational health information for health and safety professionals and consumers seeking information about the adverse effects of workplace exposures to chemical and biological agents.

Drug Information Portal

The Drug Information Portal mobile Web site provides information for prescription and over-the-counter drugs. In 2013, the resource was updated to include over 45,000 drugs.

LactMed

The LactMed mobile app provides information about drugs and other chemicals to which breastfeeding mothers may be exposed. This app is available on iPhone, iPad, and Android.

The **TOXNET mobile Web site** is a mobile-optimized version of the TOXNET databases, covering toxicology, hazardous chemicals, environmental health and related areas.

Embryo

The Embryo mobile app is an educational tool that allows users to visually explore human embryo development. The app presents 2D and 3D digitized images of the Carnegie Embryo Collection. The project is collaboration between NLM/SIS, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Louisiana State University Health Sciences Center, and the National Museum of Health & Medicine, Human Developmental Anatomy Center. This app is available on iPhone and iPad.

Disaster Information

The Disaster Information Management Research Center continued to perform its role in providing authoritative health information for disaster preparedness, response, and recovery and to assist with information needs and dissemination for specific incidents. In October/November 2012, Hurricane or Superstorm Sandy struck the Mid-Atlantic and New England coasts. Millions of people were affected in some way, such as displacement from homes and loss of heat and power, and several hospitals were evacuated due to potential or actual infrastructure issues. DIMRC created a Web page with local links as well as information about health issues. In addition, the disaster outreach listserv and Twitter were used to share information about the status of library and health facilities, provide health information and resources, and offer a forum for people to share information, ask questions and get support. DIMRC continues to provide access to

information about the storm and in March participated in a workshop, *Ports in a Storm: The Library as a Community Partner in Disaster Recovery,* in Eatontown, NJ. Staff presented on NLM resources and described the Disaster Health Information Outreach Projects that encourage partnerships between disaster organizations and libraries. In addition, many libraries were not aware that they can apply for FEMA assistance and relocation if damaged in a disaster under the Robert T. Stafford Disaster Relief and Emergency Assistance Act.

Disaster Information Outreach

The Disaster Information Specialist Program is a collaborative effort to explore and promote the role of information specialists in the provision of disaster-related information resources to the workforce and communities. In FY2013, DIMRC hosted ten meetings/conference calls as continuing education programs for librarians and members of the disaster workforce. These ongoing conference calls supplement the more formal Disaster Information Specialist curriculum and courses offered online and occasionally in-person by DIMRC. This year's conference calls featured presentations from a number of subject matter experts on disaster behavioral health, science during disasters, social media in disaster response for information dissemination and situational analysis, resilient communication systems, gaming technologies for incident command training, and the needs of persons with disabilities in disasters. In addition, presentations on the roles of librarians (by the ambulance riding librarian) and access to biomedical and public health information in disasters were included. The number of attendees averaged 60 callers per meeting, (However, we do not know if multiple people called in from one location.)

The Disaster Information Specialist course materials were transferred from the Medical Library Association's Web site to the DIMRC Web site this year and all course documents were remediated to ensure Section 508 compliance. Course registration and evaluation also was transferred to NLM.

A new and very popular webinar entitled *Packing Your Digital Go-Bag: Essential Disaster Health Information on Your Mobile Device* was developed and presented to over 300 people in two sessions, one hosted by DIMRC and the other by the NN/LM Mid Continental Region. The presentation was originally conceived and designed for a presentation at the March 2013 Public Health Preparedness Summit. Due to government travel restrictions, DIMRC's participation in that meeting was cancelled. Students may earn credit for this webinar towards their Disaster Information Specialist certificate.

At the end of FY2013, NLM awarded purchase orders to three organizations for disaster information outreach projects. The projects required an organization working in disasters (public health department, emergency management department, fire/rescue service, academia, etc.) to partner with a library to improve access and knowledge of NLM and other disaster health resources and to identify roles for librarians in disaster management. At the end of FY2012, NLM awarded seven purchase orders and the projects were completed during FY2013. These projects include assisting and training the Medical Reserve Corps of Arkansas, teaching emergency responders and public health officials to use mobile apps related to disaster health, developing situational awareness briefs, and assisting with information management issues related to disasters including family reunification, situation awareness in the hospital command center, etc. The three projects funded at the end of FY2013 include enhancing access to health and emergency preparedness information in El Paso, Texas, and training librarians, public health personnel, and emergency responders to use disaster health information tools in American Samoa and in Fort Worth, Texas.

DIMRC continued to make progress in promoting awareness and information sharing via the Disaster Outreach Listserv and the @NLM_DIMRC Twitter feed. The listserv membership grew from over 800 subscribers at the end of FY2012 to over 1100 at the end of FY2013. Many new members are part of the disaster medicine and public health workforce. During FY2013, DIMRC continued to use Twitter extensively to reach many new people and to assist other agencies, partners, and collaborators to share and disseminate information more widely. DIMRC staff prepared tweets ahead of time to be stored in a "tweet deck" and delivered automatically on a pre-determined schedule. In addition, DIMRC staff kept up with current events and activities and helped communicate timely information to a wide audience. DIMRC continued to participate in Twitter chats and tweetups such as the Social Media in Emergency Management (hashtag #smemchat). The NLM_DIMRC Twitter account tripled its number of followers from over 500 at the end of FY2012 to over 1,500 followers at the end of FY2013. Twitter allowed us to quickly and easily promote DIMRC and other NLM resources, especially after Hurricane Sandy, and to collaborate with other agencies and organizations involved in disaster medicine and public health such as the Assistant Secretary for Preparedness and Response (@PHEgov) and the National Center for Disaster Medicine and Public Health (@NCDMPH). DIMRC also joined the new HHS Social Media in Emergency Management and Public Health working group in September.

DIMRC presented at a number of meetings including teaching the *Digital Go-Bag* class at the World Congress on Disaster and Emergency Medicine (WADEM) and to the Montgomery County, Maryland Community Emergency Response Team, presenting in sessions on social media in disasters at the Pan American Regional Conference of WADEM and the Pan American Health Organization Library and Information Science conference, demonstrating disaster response tools at a Technology Petting Zoo sponsored by the DC Special Library Association Meeting, discussing DIMRC tools and

resources at a post-Hurricane Sandy library conference, and participating in a disaster drill and exercise in California. DIMRC also exhibited at several conferences on disaster medicine and emergency response.

DIMRC continued to work with the Latin American Disaster and Health Information Network (LANDHI) throughout FY2013. Projects funded at the end of FY2012 were completed, including an electronic version of the only Central American disaster management journal, a prototype geographical information system (GIS) for retrieving disaster management documents in the region, and an Android app to access the Web sites of LANDHI and the participating organizations, a new LANDHI Web site, promotional materials, and a learning management system to support network training courses. A new center, the National Office of Emergency Management of Chile, joined the network, which now spans 16 disaster and health information centers in 11 countries. The network members met in October 2012 to present projects and discuss plans for the coming year.

Approximately 100 documents were added to the Haiti Disaster Information Center Web site. Efforts continue to identify an organization in Haiti to host, support, and maintain the Web site.

Disaster Health Information Resources and Tools

Following a restructuring of the Resource Guide for Disaster Medicine and Public Health last year, the focus in FY2013 was to improve and enhance the search features and develop procedures to streamline the scanning, selection, and input of new materials into the database. Over 1,400 items were added this year. The search engine now includes Boolean searching and search filters to assist users in quickly narrowing searches by source, publication type, date and author.

Several new Web pages were developed in response to incidents including Hurricane Sandy, Hurricane Sandy Recovery, Influenza H7N9, and Coping with Disasters, Violence, and Traumatic Events (after the Newtown shooting). In addition, a page on information sources on Ethics in Disaster Medicine and Public Health was developed. Over a dozen disaster topic pages were updated.

SIS continued working with the HHS Assistant Secretary for Preparedness and Response (ASPR) on the development of an interactive, searchable *All-Hazards Plan* (previously called the *Playbook*). ASPR staff wrote the new plan to be used by HHS staff at headquarters and in the field offices. NLM turned the document into a database-driven resource that can easily be searched and provide quick and easy access to specific portions of the resource. The new interactive tool has been used in the Secretary's Operation Center (SOC) for several incidents, including Hurricane Sandy. This new tool allows staff in the SOC to quickly identify what needs to be done at each phase of an incident (before, during, and after) and who needs to do it. NLM made a number of enhancements to the tool this year and began developing methodology to integrate incident-specific annexes (e.g., hurricane, earthquake, chemical emergency) into the tool.

Under an Inter-Agency Agreement signed last year, staff from SIS and the Lister Hill Center for Biomedical Communications began working with the FDA on the development of an app for data capture of adverse effects of drugs approved via Emergency Use Authorization in Public Health Emergencies and provided guidance to FDA and its contractor on app and mobile Web site development. SIS will provide usability testing and support for the tools as they are developed throughout FY2014.

At the end of FY2013, NIEHS received funds from the NIH Director's Discretionary Fund to develop a pilot NIH Research Response project. This project will investigate mechanisms to conduct research following a disaster in a timely fashion. The project includes a survey of NIH disaster-related projects, the development of an inventory of surveys and other data collection tools that can be used in disasters, and coordinating the work of existing research networks to help conduct research in disasters. NLM will develop the public Web site for the project.

WISER and CHEMM

The major enhancement for the WISER (Wireless Information System for Emergency Responders) and CHEMM (Chemical Hazards Emergency Medical Management) emergency response tools was to integrate the content of both resources into a single app that is available for iOS and Android devices. This approach was chosen in response to user feedback. Emergency responders indicated that they would prefer one app as they would not remember multiple apps, particularly in a large-scale disaster. They preferred an app that they use on a regular basis and with which they are already familiar. So WISER's large user base (nearly 500,000 downloads) was leveraged and the CHEMM content was integrated in a manner that is easy/intuitive for users (via role-based navigation). The addition of CHEMM content expands WISER's user base to include hospital providers and emergency planners (in addition to the original first responder, hazmat specialist, and EMS roles), thus increasing its diversity and potential reach.

In addition, DIMRC worked with the Department of Transportation's Pipeline and Hazardous Materials Safety Administration (PHMSA) to both incorporate the new *Emergency Response Guidebook* (ERG) 2012 into WISER and also to develop an ERG 2012 app for iOS and Android. DOT requested NLM's assistance due to our expertise with app development.

A new contract was awarded in FY2013 to update and enhance the CHEMM content.

REMM

The Radiation Emergency Medical Management (REMM) resource was enhanced in several areas this year: new videos, illustrations, photos, and graphs were added to the REMM Multimedia Library. These additional assets are made available to health care professionals to teach radiation concepts and train in radiation medical response. The key topic pages, such as nuclear detonation and nuclear power plant incidents, were updated significantly with new content and references. Key content pages were created for specific audience groups, such as the medical examiners/coroners, and public health emergency researchers.

Development began on redesigning the Mobile REMM on iPhone, Android, and BlackBerry platforms. New interactive tools are being developed, including the radiation unit conversion tool and the Scarce Resources Triage Tool after a Nuclear Detonation. A redesign of the Acute Radiation Syndrome management tool was also implemented in 2013.

Health Information Technology Research and Development for Disaster Management

DIMRC continued to coordinate and conduct a number of informatics and communications research and development projects that were initiated several years ago by 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. NLM is now seeking additional partners to broaden the reach of several of these projects conducted by SIS, OCCS, and LHNCBC.

DIMRC continues to develop, enhance, and test the Virtual Hospital Incident Command System (HICS) training system. The second Emergency Operation Center (EOC) virtual exercise was conducted at Suburban Hospital. The exercise simulated a large-scale disaster that severely damaged the hospital's infrastructure and, for the first-time, the staff practiced the decision-making and organizational processes needed for a full evacuation of the hospital facility. Walter Reed National Military Medical Center staff also participated along with observers from many emergency and health management organizations including FEMA and NCDMPH. The trainees indicated in the post-exercise review ("hot wash") that the virtual learning environment was a very effective way for them to train and practice HICS. The project was highlighted in an NLM In-Focus article and a poster was presented at the Learning in Disaster Health workshop. A conceptual design was developed (to be implemented in FY2014) for a single-user tool to introduce trainees to the virtual training platform and to provide individual practice.

DIMRC also continued to develop the MARS Emergency Radio system to improve and simplify the communications tools between the users and the radio operators and the station management tools for the operators. The system was installed and tested by radio operators in Kent County, Maryland, and they are testing the system as a means to provide emergency communication services to the local emergency management organizations. In addition, the system was tested at several events at the Maryland Emergency Management Agency, NIH and other radio clubs, an amateur radio field day, and at the BHEPP exercise. These tests were successful and indicated that a more compact design is needed for rapid deployment.

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 Institutes (including a new project on disaster research response with NIEHS, the NIH Biodefense Research Coordinating Committee, and the NIH Disaster Interest Group), the Department of Transportation, the Food and Drug Administration, and the Institute of Medicine's Forum for Medical and Public Health Preparedness for Catastrophic Events and other agencies.

Outreach Initiatives

Community Outreach

In FY2013, SIS staff continued its outreach efforts working with Black Greek Letter Organizations. These organizations are involved actively in local outreach efforts to combat the health disparities that adversely affect minorities. Staff worked with members of Kappa Alpha Psi Fraternity to promote health literacy and health awareness.

SIS funded three innovative outreach projects in information dissemination for family and women's health. Public libraries are strategic partners in increasing the awareness and utilization of NLM and National Institutes of Health (NIH) health resources, and meeting NLM long range goals of health literacy, informing citizens, and reducing health disparities. All projects addressed family health and target women as the main information gatherer and health decision influencer in the family. Awardees are Petersburg Public Library System (VA), Pima County Public Library (AZ) and Forsyth County Public Library (NC.)

Outreach to Minority Populations

The goal of the Environmental Health Information Partnership (EnHIP) is to enhance the capacity of minority serving academic institutions to reduce health disparities through the access, use and delivery of environmental health information on the campuses of HBCUs, tribal colleges and minority serving institutions and in their

communities. EnHIP, now in its 22nd year, has 22 partnering institutions including 14 HBCUs, three tribal colleges, three Hispanic-serving institutions, one Alaskan university and one community college. In March, members met for a stimulating program entitled "3Rs: Recruitment/Retention, Research, and Relationships" featuring a keynote address by Shirley Malcolm, PhD, Association for the Advancement of Science, on broadening the talent base for STEM. In addition, Keith Gregory, Hampton University Proton Therapy Institute, Dr. Keith Norris, Charles Drew University of Medicine and Science and Dr. Emma Fernandez-Repollet, University of Puerto Rico Medical Sciences Campus presented and discussed innovative programs to recruit and retain minority students, promote research by faculty and students and develop entrepreneurial ventures based on campus research. Dr. Ann Barbre, Xavier University, chaired the meeting and discussed ways to further relationships and communication among members. Four small outreach projects were funded with EnHIP partner institutions in FY2013.

SIS continued its efforts to increase the awareness of and use of NLM online resources by minority health professionals. Hands on and online training, presentations, and demonstrations were conducted at the Student National Medical Association Annual Medical Education Conference, National Medical Association Annual Convention and Scientific Assembly, Black Nurses Association annual conference, National Association of Hispanic Nurses conference, National Indian Health Board, USET–United South and Eastern Tribes conference, Conference on the History of African Americans in the Medical Professions (CHAAMP) conference and the Sixth Annual National Conference on Health Disparities.

Staff participated in the Trans-NIH AI/AN workgroup's workshop, *Native American Elders: Fostering Interagency Collaborations to Support Effective Outreach* with the Indian Health Service and Department on Aging. The meeting addressed needs and challenges of AI/AN elders, successful partnerships and dissemination models, and opportunities for collaboration. SIS shared its experience in working with AI/AN populations in community-based information outreach projects.

In July, SIS hosted a visit by Washington Internships for Native Students (WINS) interns. The group toured NLM and the *Native Voices* exhibition and learned about NLM resources & services.

Outreach to Librarians

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

NLM was a sponsor of the Eighth National Conference of African American Librarians and provided support for the health information program at the conference. NLM supported several programs, including Women's Health Issues for Diverse Communities presented by Vivian W. Pinn, MD, Senior Scientist Emerita, Fogarty International Center, National Institutes of Health; *Clinical Trials Today: Do We Still Need To Fear Tuskegee?*, presented by James H. Powell, MD, President, Strategic Medical Associates; and The Homecoming: A Time for Celebration or Crisis presented by Evelyn L. Lewis, MD, MA FAAFP, President and Chief Executive Officer of Evelyn Lewis International, a health care strategy and consulting service with a focus on multicultural and diverse populations. NLM also sponsored speaker Goulda Downer, PhD, RD, LN, CNS, Principal Investigator and Project Director, National Minority AIDS Education and Training Center (NMAETC), Howard University, College of Medicine. Dr. Downer's presentation entitled, *Taking Responsibility for Your Good Health: Your Actions and Choices Make All the Difference* generated tremendous response and was the catalyst for a health and wellness information project with the Black Caucus of the American Library Association.

Special Population Web Sites

In FY2013, SIS initiated a redesign of the **American Indian Health Web Portal** (AIH) to enhance the content organization and graphical design of the portal. SIS worked with the Center for the Study of Asian American Health (CSSAH) at the NYU Langone School of Medicine on a needs assessment to address the current and future health information needs of health providers, researchers and others providing services to Asian Americans, information delivery methods, and the need for non-English materials. Survey findings are being used to focus future directions of the **Asian American Health Web Portal**.

Arctic Health was re-launched in May with an updated interface and content sections and a new URL http://arctichealth.nlm.nih.gov. The portal features new images from the University of Alaska Anchorage (UAA) and uses colors and textures reflective of the Arctic environment. Emphasis was placed on the use of stories to disseminate indigenous knowledge as well as on the topic of climate change which is very important in the arctic.

Women's Health Resources (WHR,) a collaboration with the Office of Research of Women's Health (ORWH), NIH, continued to expand research dissemination through the Web portal and social media. A yearlong evaluation with focus groups and usability testing of the portal was completed. A redesign based on the findings is planned. Based on the study, SIS staff presented the poster, *What Women Want: Re-envisioning the Women's Health*

Resources through User Needs and Usability Testing at the MLA conference.

In FY2013, SIS worked with ORWH in an outreach effort to increase the capacity of sex differences research by university researchers and increase knowledge and use of this research. Nine projects with an intended audience of undergraduate, graduate and professional students, faculty and health professionals were awarded. Awardees included: Arizona Health Sciences Library, University of Arizona; Earl S. Richardson Library, Morgan State University; Hardin Library for Health Sciences, University of Iowa; Lamar Soutter Library, University of Massachusetts Medical School; Lister Hill Library of the Health Sciences, University of Alabama; Medical University of South Carolina Library; Oviatt Library, California State University at Northridge; Spencer S. Eccles Health Sciences Library, University of Utah; and Health Science Center Libraries, University of Florida.

Improvements were made to the **Refugee Health Information Network (RHIN)** a partnership of the Association of Refugee Health Coordinators, NLM and the Center for Public Service Communication (CPSC) RHIN provides 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. In FY2013, SIS improved the content management system to make it easier to update records. Also added was a "Group" function, so all formats and languages for a given title and author are listed under one unique ID. This allows users to view all available formats and languages at one time. Users submitting content for inclusion in RHIN can now define their uploads with Creative Commons, Public Domain, and Copyright distinctions. SIS also added Healthy Roads Media, King County, and Ready.gov documents in all available formats to the database.

The Outreach and Special Populations Branch (OSPB) provides distance education, listservs, and social media to various populations, including professional organizations, organizations serving minority communities, and consumers. OSPB released an enhanced version of its Web site (http://sis.nlm.nih.gov/outreach.html) to improve access to many specific population resources and programs. In addition, OSPB launched a new communications strategy to inform the librarians, specific populations, and outreach and community health organizations on new resources from NLM and HHS, funding opportunities across the HHS and the federal government, and new materials on the Web page. The strategy includes use of Twitter, @NLM_OSP: https://twitter.com/NLM_OSP, and a listserv, NLM OSP.

Social Media

SIS continued its use of social media for promotion and outreach this year. Staff continued to maintain and improve Facebook accounts for several SIS resources, including NLM 4Caregivers, Women's Health Resources and AIDSInfo. The social media team manages several Twitter feeds to promote and update users on resources and programs. A new Twitter feed, @NLM_HIVplus50, was introduced to provide news, updates and resources on HIV/AIDS for the aging population.

OSPB hosted four community health student interns from the University of Maryland School of Public Health who worked with the social media team on Twitter, the Tweetbank, Hootsuite and Facebook. They created communications plans, original tweets, social media metrics, and workflow schemas along with developing a local outreach effort to promote NLM resources. In March, the social media team hosted a University of Michigan Alternative Spring Break Student, who developed a Twitter/social media plan for the RHIN database. In June, SIS welcomed an Adopt-A-School student from Wilson High School, DC, for a summer internship to assist with NLM4Caregivers and other social media efforts.

K-12 Initiatives and Resources

NLM is partnering with King Drew Medical Magnet High School in Los Angeles, Kotzebue High School, Alaska, and Farrington High School, Hawaii, to deliver a distance learning program via synchronous videoconferencing and collaboration technologies using Internet2. The program includes curriculum development, student evaluation, and two-way video. NLM kicked off the Distance Learning Project this year in August. The theme for the 2012-2013 sessions was *Science, Technology and Math in Healthcare*, with presentations by NIH personnel and others. In July, SIS hosted a visit for six high school teachers and IT staff from participating schools. While at NLM, they met with various staff to discuss and learn more about social media, apps, the Native Voices exhibition, K-12 resources, the National Network of Libraries of Medicine, Profiles in Science, History of Medicine Division exhibitions and resources, virtual world programs and other resources that would be meaningful to their work.

The **South Carolina Teen Health Leadership Program** teaches minority high school students to navigate and assess health information on the Internet using MedlinePlus as well as other NLM and Medical University of South Carolina (MUSC) resources. In addition, the program promotes health careers, promotes healthy behaviors, and develops leadership skills through student-community engagement. In April SIS hosted the students from St. John's High School, John's Island, SC who participated in the program. The students presented a program on autism awareness, their outreach topic for the year, including a PSA developed on the topic. In June, SIS hosted Deborah Williamson, Associate Dean for Practice, MUSC College of Nursing, who provided a more in-depth description of the origins of the program, the culture and environment of John's Island, and the outreach done by MUSC staff with K-12 students in Charleston. A

community health student intern from the University of Maryland School of Public Health is working on an qualitative evaluation of the Teen Health Leadership Program.

SIS participates in the **NLM Adopt-a-School Program**, a partnership with the NLM Office of the Director. The program provides training in the use of NLM's consumer health databases, summer internships for students, enrichment of library resources including the donation of books and periodicals, guest lectures, and more. SIS coordinated summer internships for five students from Woodrow Wilson Senior High School in Washington, DC and Charles H. Flowers High School in Maryland. Hosts included Specialized Information Services (SIS), Library Operations (LO), Office of the Director (OCPL), Lister Hill Center (LHC) and Office of Administrative and Management Analysis Services (OAMAS.) Students worked from June-August and submitted papers about their experience at the NLM.

Tox Town was enhanced with a new "Drought" location and hydraulic fracturing was added to the farm scene. Methane was added as a new chemical. In FY2013, SIS endeavored to increase its promotional efforts for Tox Town and its corresponding curriculum units. The site was added to several organizations' Web sites, such as NIEHS Kids' pages and the National Science Teachers Association (NSTA) Web site. It was featured along with ToxMystery during HHS Earth Day activities at HHS headquarters in DC. Staff began a "Tox Town Chemical of the Week" tweet.

NLM SIS Twitter @NLM_SIS
Tox Town Chemical of the Week #TTCoW: Persistent Organic Pollutants (POPs). Learn more: http://toxtown.nlm.nih.gov/text_version/chemicals.php?id=86

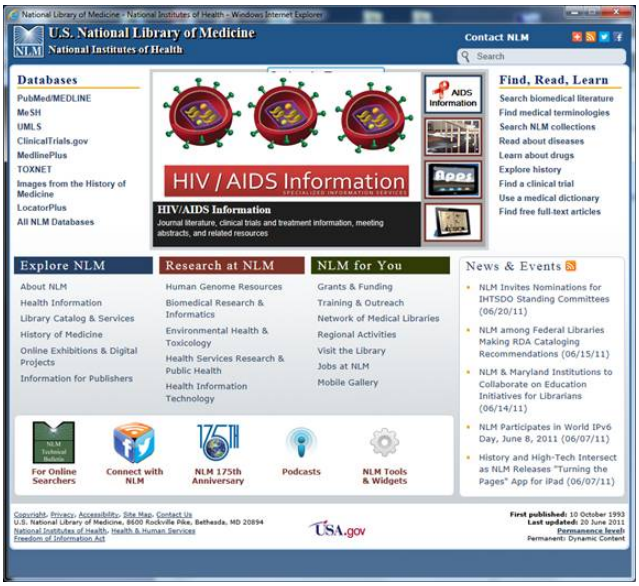
The Environmental Health Student Portal is a Web site for middle school students and teachers, developed in FY2010. In FY2012, SIS completed a usability study of the portal, which produced valuable information for improving the site and led to redesign efforts, which began in FY2013 and will continue into FY2014. The redesign involves simplifying site navigation, increasing readability of texts introducing various topics and subtopics, and improving the search function.

Content was added and a new lesson plan created for **GeneEd**, a Web site that provides a single repository to Genetics research, study guides, lesson plans, experiments and activities for teachers and students in grades 6-12. The resource was promoted at exhibits and through the K-12 newsletter.

The K-12 team worked with an NLM Associate Fellow on a project to develop two short videos to promote NLMs K-12 resources to educators. One video is a live action feature with interviews with teachers discussing NLMs genetics resources (GeneEd, Harry Potter's World and Genetics Home Reference); the other, an animated video, features resources for middle school students. As additional NLM resources and programs relevant to the K-12 population (educators and students) are identified they are added to the NLM K-12 Science and Health Education Web page maintained by SIS (http://sis.nlm.nih.gov/outreach/k12.html).

Native Voices: Native Peoples' Concepts of Health and Illness includes a touring and a virtual version of an exhibit that explores the interconnectedness of wellness, illness, and cultural life for Native Americans, Alaska Natives, and Native Hawaiians. SIS developed educational activities and created lesson plans around the exhibit for students in grades 6-8. The four lessons use exhibit videos to teach students about similarities and differences between Indigenous and Western perspectives on land, community, the environment, health and healing. Students also learn about the impact of Western expansion and industrial development on Native Peoples' environment, culture, and health.

HIV/AIDS



We continue to use many avenues of communication to reach our diverse user communities. The above graphic features an image of the human immunodeficiency virus specifically designed for use on consumer materials for AIDS*info*.

AIDS Community Information Outreach Program

NLM's AIDS Community Information Outreach Program (ACIOP), 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, resource development, and equipment acquisition. In addition to a program evaluation, a workshop was held with ACIOP awardees and leaders in the field of HIV/AIDS information access regarding emerging trends in technology and external changes that might justify program changes and reprioritization of funding. Basedon recommendations received, result, NLM modified the FY2013 Request for Proposal (RFP), significantly changed requirements for ACIOP quarterly and final reports , conducted webinars for awardees on NLM HIV/AIDS resources and evaluation assistance, and implemented a blog for communication between and among ACIOP awardees.

Among the 13 new HIV/AIDS Community Information Outreach Projects funded in late FY2012 were five projects that focused on the use of mobile and smartphone technology as well social media platforms to increase the awareness and utilization of HIV/AIDS information resources. In September 2013, NLM funded seven new projects that build upon previous efforts. The Hispanic Communications Network, a social marketing agency that focuses on disseminating health information to the Hispanic community as well as produces Spanish language health information for Spanish radio in the US, will leverage its national media network, online resources and partnerships with Hispanic–serving HIV/AIDS organizations to produce and syndicate culturally and linguistically appropriate messages to increase awareness and access to *info*SIDA. They produced two Spanish PSAs that were distributed across HCN's national network of Spanish–speaking radio stations and aired throughout the month of June. They also invited a Spanish-speaking librarian from NLM to be on the live radio program, *Bienvenidos a América* in May. The program was recorded and is available on the HCN and BAA sites. Another project includes *Comunidades Unidas*, which will deploy mobile technology by *promotoras* and expand their text messaging system and social media campaigns. Similarly, *Project HIVE I* in Washington, DC will develop a "health information virtual exchange" that will serve as a one-stop HIV/AIDS resource for African Americans' growing use of smartphone technology.

NLM AIDS portal

Close to 200 new links were added to the SIS AIDS portal Web site, including 58 Spanish-language resources. A new topic area was developed for mobile resources. Staff completed updates to the Spanish version of the *HIV/AIDS Resources from the National Library of Medicine* workbook. Both the Spanish and updated English versions are available on the AIDS portal: http://sis.nlm.nih.gov/outreach/hivaidsworkbook.pdf.

*AIDS*info/info*SIDA*

This year there was a major focus on mobile dissemination –a big step towards making HIV/AIDS treatment information easily available at the point of care. All five HHS treatment guidelines managed by AIDS*info* were made available in a new mobile optimized format. All the content including recommendations, tables, and text are now easily accessible from mobile devices. The content management system (CMS) was modified to support this change and now all formats of the guidelines are produced from the CMS, eliminating most potential versioning problems.

Other content was updated during the year as well. All the HIV/AIDS treatment fact sheets, both English and Spanish language) were revised and reorganized. These fact sheets are intended for use by consumers and the new HTML versions make them readily searchable, fully 508 compliant and readable on mobile devices. There are also printer-friendly versions available.

Social media outreach continued to grow during this year. By the end of FY2013, the AIDS*info* Twitter account had 5,793 followers and the Facebook page had 3,417 fans. There is a combined total of 8,147 subscribers to the e-newsletter (English and Spanish language editions).

Evaluation Initiatives

In FY2013, SIS continued to use The American Customer Satisfaction Index (ACSI) to obtain user input on AIDS*info* and *info*SIDA sites. AIDS*info* has consistently scored above 80, which is much higher than most federal government Web sites.

In FY2013, we continued working with a team from Columbia University on data collection and analysis for the evaluation of the AIDS Community Information Outreach Program. Jointly, we created a project report, which was presented to the NLM Board of Regents, and wrote a journal article, currently under review for publication in a peer-reviewed journal. The article summarizes results of the evaluation study and lessons learned during the project.

As an effort to continue evaluating its programs and Web sites, SIS started a qualitative evaluation of the South Carolina Teen Health Leadership Program (THLP), conducted with Medical University of South Carolina (MUSC). SIS recruited a full-time student intern from the University of Maryland at College Park community health program, developed evaluation instruments, and worked with MUSC on reaching out to THLP graduates interested in participating in the study. We plan to continue the evaluation into FY2014.

SIS also laid out a protocol for conducting a focus group evaluation of the Refugee Health Information Network (RHIN) Web site. RHIN is the site that has originally been developed to serve information needs of healthcare providers, working with refugees and asylees.

SIS is currently considering expanding the site's focus to include non-refugee non-English, non-Spanish speaking immigrant groups in the US. The objective of the focus group protocol was to assess RHIN current and potential users' perceptions associated with such change. The evaluation will continue into FY2014.

Research

SIS participated in a number of research activities in the field of health information outreach, consumer health informatics, and science education. We conducted a study of public librarians' treatment of consumer health information queries, focusing on print and Web resources recommended to patrons. We also interviewed public librarians about their responses to a hypothetical scenario, in which a patron asked for resources, explaining possible connection between childhood vaccination and autism. Both studies were conducted in collaboration with the school of library and information studies at the University of Wisconsin, Madison. The study of general consumer health information queries in public libraries resulted in a manuscript, accepted for publication at the *Journal of Consumer Health on the Internet*. We also worked on the manuscript, describing the findings of the second (vaccination query) study.

SIS Photo Gallery

Dr. Brenda Manuelito (l) and Dr. Carmella Rodriguez, co-founders of NDigiDreams Inc., a training company specializing in instructional technology and digital storytelling. Presenting their work with NLM at the NIH 204 American Indian and Alaska Native Heritage Month.

One-on-one at the Health Living and Learning Center, Petersburg Public Library.

Pat Higgenbottom, Associate Director for Public Services for Lister Hill Library of the Health Sciences, University of Alabama, Birmingham, and Noruwa Agho, UAB School of Public Health intern, at the afternoon tea. To do more personal outreach, the library staff worked with liaisons and interns from the Schools of Nursing and Public Health. They held an educational afternoon tea at the library with faculty, staff, and students, and promoted the portal during a meeting of the Commission on the Status of Women.

Promotional poster for services available at the Forsyth County Public Library system, North Carolina.

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 US Congress in 1968, is an intramural research and development division of the National Library of Medicine (NLM). LHNCBC seeks to improve access to high-quality biomedical information for people around the world. It leads programs aimed at creating and improving biomedical communications systems, methods, technologies, and networks and enhancing the sharing and use of information among health professionals, patients, and the general public. The development of next generation electronic health records (EHRs) to facilitate patient-centric care, clinical research, and public health is an important focus of the LHNCBC as well as an area of emphasis in the *NLM Long Range Plan 2006−2016*.

The LHNCBC research staff is drawn from many disciplines including medicine, computer science, library and information science, linguistics, engineering, and education. Teams of people from a variety of backgrounds conduct research that often involves collaborating with other NLM divisions, institutes at NIH, organizations within the Department of Health and Human Services, and academic and industry partners.

LHNCBC is organized into five major components: the Cognitive Science Branch (CgSB); the Communications Engineering Branch (CEB); the Computer Science Branch (CSB); the 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. News and information about LHNCBC research activities are available at http://lhncbc.nlm.nih.gov/.

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

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

NLM Personal Health Record

The NLM Personal Health Record (PHR) is a Web-based tool designed to help consumers track their health information. The goals of the PHR are to:

- Help consumers manage and understand their healthcare problems.
- Facilitate federal goals for clinical data interchange using national vocabulary standards.
- Determine whether using personal health records can improve adherence to preventive-care recommendations and improve consumer health.

Consumers can use the main PHR page to enter key health information, including medical conditions, surgeries, medications, allergies, and immunizations. They can also enter dates for prescription refills and medical appointments and doctors' contact information and record questions they want to ask their doctor. From the main page, they can find out about their own health issues by consulting MedlinePlus and other trusted resources. Consumers can enter data for lab results, radiology reports, and other screening and diagnostic procedures. In addition, they can track measures of wellness including mood, diet, sleep, and exercise, as well as disease-specific parameters such as episodes of asthma or frequency of seizures.

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 (for example, Logical Observation Identifiers Names and Codes (LOINC), RxNorm, and Systematized Nomenclature of Medicine−Clinical Terms (SNOMED CT)). By using vocabulary standards, the PHR can automatically remind consumers about the preventive care and healthy behaviors specific to them, based on the data they entered.

In FY2013, LHNCBC researchers and developers continued to improve the PHR. We launched a demonstration site where people can test out the system without actually storing their data.

On the technical side, we updated JQuery from version 1.6 to 1.9 and implemented 1.9s datepicker, a specialized calendar function. We also updated software to Ruby 2.0 and Rails 4.0. We continued working on ways for users to import into the PHR their health data from continuity of care documents (CCD) – which hospitals and doctors will issue to their patients based on regulations to ensure the meaningful use of electronic health records (EHRs).

On the content side, we revised the content and wording of multiple preventive-care reminder rules, including those for lipid screening, bone density, cervical cancer screening, and breast cancer screening. We also added more consumer-oriented terms and lists of synonyms for medical conditions and surgeries, including common orthopedic and cosmetic surgery terms.

We conducted a formal usability study of the PHR and implemented changes based on the results that

make the site easier to use. To simplify data entry, we revised the allergy and immunization tables, and we added a section on the main PHR page where users can view previously entered data and enter new data points for screening tests. We redesigned the health-panel organization and display and the health reminders, and we added visual cues to the main PHR page to ensure that new and unread reminders are displayed more prominently. We added tooltips, such as question-mark icons that lead to help messages, throughout the user interface, and we began developing methods for collecting and reporting PHR usage statistics. We are also exploring using the PHR to collect patient-reported data for clinical trials.

This project addresses the long-standing NLM interest in EMR systems and delivery of healthcare information to consumers and is closely aligned with the NLM strategic plan. It uses the nationally mandated vocabulary standards that NLM has supported, and it will provide another consumer entry point to NLM's rich trove of patient-oriented data.

Using Drug Databases to Assess Prescribing Practices and Continuity of Care

Getting an accurate medication history for Emergency Department (ED) patients is important for their emergency care, especially since a significant proportion of ED visits are related to adverse events from prescription medications. Gathering such information from patients is time-consuming, expensive, and sometimes impossible (such as when a patient is unconscious), and patient-provided medication histories are often incomplete. Since 2009, Suburban Hospital's ED has used Surescripts, a consortium of major pharmacy benefit managers (PBM), to provide an electronic summary of the prescriptions their patients filled over the past year.

LHNCBC created an electronic messaging interface engine, based on *Mirth* (an open-source Health Level Seven, Inc. (HL7) interface program), which linked with Surescripts and delivered the prescription records for patients who had checked in for ED care at Suburban. Before the system went live, Suburban Hospital collected both Surescripts data and patient-provided histories. LHNCBC researchers obtained this information in a de-identified form, and then compared the two sources of information. We found that Surescripts information, when available, significantly augmented the history obtained in patient interviews by 28 percent (adding 1.1 prescription drugs per patient) and covered a high proportion (88 percent) of a patient's current medications. These findings were published in the September 2013 issue of the *Annals of Emergency Medicine*.

The concise prescription-dispensing history report that we developed (based on the Surescripts data) is now routinely provided for patient care in the Suburban ED. Clinicians there report that the full-year prescription history is also helpful in identifying potential problems of drug compliance, drug-seeking behavior, and duplicative prescriptions. A brief follow-up study by the ED pharmacists showed that patients were indeed taking 85 percent of the medications they had not mentioned during the patient interviews.

We're also using prescription databases to study adverse drug events due to drug to drug interactions. With a commercial insurance dataset from the Washington, DC, metropolitan area covering 2 million patients over one year, we're studying the incidence of potentially dangerous drug to drug interactions and whether adverse reactions have actually occurred (for example, has a patient returned to the ED?). We're evaluating the alerts generated by several commercial and publicly available drug to drug interaction knowledge databases by applying them to this prescription dataset.

EMR Database Research and Natural Language System Development

In FY2013, we continued working with the sixth update of the MIMIC-II electronic medical record (EMR) dataset. These big datasets are helping us to answer clinical research questions. For example, we're using the de-identified MIMIC-II data under a restricted-use memorandum of understanding to conduct retrospective clinical studies on the:

- Significance of obesity and as a risk for mortality in intensive care units (ICUs).
- Interactions between feeding practices and blood transfusions in premature babies with necrotizing enterocolitis.
- Significance of vitamin B12 levels in ICU mortality and post-discharge survival.

Our analysis has shown that in the ICU, overweight and obese patients have a survival advantage over normal-weight patients. Our findings, published in the *Critical Care* journal, were followed by a letter to the editor of the *Journal of the American Medical Association* (*JAMA*). Manuscripts describing the results of the necrotizing enterocolitis and the vitamin B12 studies were submitted for publication.

In line with the NLM mission to facilitate access to health information resources, we continue to serve as a mirror site for PhysioNet, a very large (4.3 TB) and widely-used database of physiologic waveform tracings gathered from healthcare institutions worldwide by the same Massachusetts Institute of Technology researchers who developed MIMIC-II. These waveform data are collected by sensors attached to the patient and are used to monitor different aspects of a patient's clinical status, including heart rate and blood pressure, respiratory status, intracranial pressure, and sleep.

We continue to update the MIMIC-II waveform data and provide access to this collection. We've also continued to develop information retrieval and natural language processing (NLP) techniques for extracting important clinical variables from the rich narrative text in MIMIC-II. For example, in FY2013 we developed and

published a method for extracting maternal data from a newborn's clinical notes. This is important because maternal history directly impacts newborn care but is typically only present in the newborn's record as narrative text.

Biomedical Imaging, Multimedia, and 3D Imaging

The objectives of this research area are to:

- Build advanced imaging tools for biomedical research;
- Create image-based tools for medical training and assessment.
- Develop multimedia image/text databases that accentuate 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

The American Society for Colposcopy and Cervical Pathology (ASCCP) has been using one of our image-based systems, the Teaching Tool, to assess the knowledge and skills of colposcopy professionals. More than 100 resident programs in Obstetrics/Gynecology and Family Practice at more than 95 universities and other premier institutions such as the Mayo Clinic have been using the tool. Since we first released the Teaching Tool in May 2010, these programs have administered more than 1,700 individual online exams of two types: the Residents' Assessment of Competency in Colposcopy Exam (RACCE) and the Colposcopy Mentorship Program (CMP) exam. In 2013, we worked with the ASCCP to update much of the content of these exams so that they're now consistent with the new guidelines for managing abnormal screening tests and cervical intraepithelial neoplasia (CIN). Planning is under way for a more advanced colposcopy resident exam that will also to be administered through the Teaching Tool.

Our Boundary Marking Tool, another imaging program, continued to be used by National Cancer Institute (NCI) staff and their collaborators around the world, including in Senegal, Costa Rica, Nigeria, the Netherlands, Spain, and the University of Oklahoma Health Sciences Center. With the tool, researchers have collected and annotated colposcopy images for biopsy studies and created a worldwide database for cervical cancer research.

We also continued research into methods for the computerized analysis and classification of cervical tissue using images collected and annotated by pathologists at the University of Oklahoma with the Communications Engineering Branch (CEB) Histology Image Assistant (CHIA) (formerly called the CEB Virtual Microscope). This work, with collaborators at the Missouri University of Science and Technology, includes applying our algorithm to carry out nuclei segmentation within the epithelial regions of the tissue and automated classification of the epithelium into classes of normal or various grades of abnormal (CIN1, CIN2, and CIN3).

In 2013, we successfully installed and operated the technology we developed to rapidly locate segments of epithelial tissue within large histology images of the uterine cervix. Developed in collaboration with researchers at Texas Tech University, the technique can reliably locate the epithelium one or two orders of magnitude faster than previously reported when the images are stored in formats commonly used for very large images. The method first uses compression information stored in the file to roughly separate tissue regions from background and then uses graphical processing unit (GPU) computation to classify the tissue regions into epithelial and nonepithelial tissue, at a speed 1,500 times faster than previously reported in the literature.

We also collaborated with researchers at Texas Tech University to integrate our advanced algorithms for histology image analysis and tissue classification into a Web-accessible system. Additional collaboration with academic groups included work with researchers at Lehigh University on classifying cervical disease based on comparing images of patients' cervixes with images in a database containing the "ground truth" classification (that is, a classification validated through follow-up) of these images. This way, we can compare the classification performed by the algorithm with the classification determined by human experts.

Content-Based Image Retrieval (CBIR)

Content-Based Image Retrieval (CBIR) is an active research area in the imaging research community since many of its tools and techniques find application in systems for image indexing, search, and retrieval. Goals of this research are to find images in repositories or the published literature that are visually or semantically similar to an image or text query. For example, one chest X-ray might be visually similar to another, but semantic similarity lies in finding another chest X-ray with the same lung disease.

We have developed several practical systems and tools that rely on CBIR research. For instance, our Open-i system, described more fully later in this chapter, allows users to access more than 1.3 million figures from medical journals including photographs, clinical images, charts, and other illustrations. People can sort search results based on different types of images, starting with "regular" and graphical images. Graphical images are further categorized as diagrams, statistical figures, flow charts, and tables. Regular images are further categorized as X-ray, ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and other modalities. We use more than 15 image modalities to classify the images with a support vector machine (SVM)–based framework. These modalities are image features, such as color, texture, and shape. We successfully used our modality classification

system during international ImageCLEFmed competitions, and we incorporated it into Open-*i*.

To help people find the images they need more quickly, we also extract specific regions of interest within images. These regions are often highlighted in publications by "markers" or pointers – such as arrows and symbols; we extract these markers first. We've improved the performance of algorithms we developed for detecting arrows using image-layout analysis.

Authors place arrows on figures to highlight important regions. They can be of different colors, but they're usually in contrast with the background. An automatic arrow-detection algorithm needs to be able to detect these arrows without prior knowledge about their color. Once we have the arrow, we can correlate concepts in the figure caption (text) with relevant regions of interest on the figure, which can aid in improved image-retrieval performance. We're developing additional algorithms to retrieve images that are more relevant to the query, whether the query is submitted as text, photo, or a combination. In order to retrieve relevant images, we first determine how the visual data correspond to the concepts in the query text. Using a method that divides the image into tiles, or image patches, we group each image patch with other similar-appearing patches across all images in the database. We develop a correspondence between representative image patches selected, frequently occurring groups, and key biomedical concepts in the accompanying text. We then apply machine-learning algorithms to extend this labeling to all patches in various groups so that every image patch has a text label associated with it. Because image patches are derived from images, every image is now transformed from a pixel-based representation into an image document where patches are replaced with these group labels.

The advantage of this approach is that not only are we able to map text queries to visual data, we're also able to apply fast, traditional, text-based information-retrieval techniques to image retrieval. We're working toward extending this idea to text-phrase retrieval strategies to find images with relevant local regions of interest. Recent advances in this area have led to a retrieval relevance of 75 percent for image retrieval using text queries such as, "Find lung CT images with ground-glass opacity."

We use spatial layout of pixel intensities within the image to eliminate regions that are not likely to be arrows, then we apply structural information about the arrow shape to identify candidate arrow regions. We previously used Markov Random Field models to recognize arrow-type pointers with a precision of 85 percent and recall of 82 percent. Our new algorithmic approach has improved our precision to 94 percent and recall to 87 percent. Furthermore, our algorithm automatically detects whether the arrow is of a lighter or darker color compared with the background, which enables us to successfully apply it to a wide variety of images.

We're also developing a correlation between regions of interest in images and key biomedical concepts that appear in neighboring text, such as figure captions or other text describing the image content. Image features used to index the entire image may aggregate the details in specific regions.

Another example of the role of CBIR in our work is in our development of CervigramFinder, a research tool that automatically indexes and enables the retrieval of uterine cervix images (cervigrams) by shape, color, and texture features. Being able to search efficiently by image features is a significant step toward locating records in large databases of cervigrams and patient data, such as NCI's Guanacaste and ALTS (ASCUS/LSIL Triage Study) databases containing a total of 100,000 cervigrams. We've made advances in this area by developing algorithms that use values from several fields from the patient record, such as the woman's age, HPV-infection history, and prior sexual history, along with image features from the colposcopic exams. We're using these data to develop a model for predicting the likelihood that a patient will progress to more severe forms of HPV-based uterine cervical infections, including precancerous cells.

CBIR is also allowing us to improve the use of chest X-rays in an automated approach to detecting tuberculosis and other pulmonary diseases, which could be very useful in resource-poor countries. We've developed algorithms to automatically detect ribs, aorta, and other structures and to segment lung areas. Research continues into extracting texture features to classify lungs as normal vs. abnormal using SVM classifiers.

Other areas of our research include using distributed computing and GPUs for computer-intensive CBIR tasks, especially image segmentation. Through our collaboration with Texas Tech University, for example, we developed a method that uses GPU processing power for interactively following challenging object boundaries, such as the separation between the epithelial and nonepithelial tissue in histology slides of the uterine cervix. To support early detection and improve health care outcomes for people with cervical cancer, we plan to use these segmented epithelial regions to train doctors to detect various stages of pre-cancer.

Interactive Publications

Recognizing the increasing use of multimedia in scientific work, this project investigates and develops models for highly interactive multimedia documents that could transform the next generation of publishing in biomedicine. The project focuses on the standards, formats, and authoring and reading tools necessary for creating and using interactive publications that, in addition to text, contain media objects relevant to biomedical research and clinical practice, such as video, audio, bitmapped images, interactive tables and graphs, and clinical DICOM (Digital Imaging and Communications in

Medicine) image formats for X-rays, CT scans, MRIs, and ultrasounds.

We've developed interactive publications containing these data types and tools for viewing, analyzing, and writing them. The tools, Panorama and Forge, are analogous to Adobe's Acrobat Reader and Professional for PDF documents. Panorama, used for viewing and analyzing these publications, was 1 of 9 semifinalists out of 70 entrants in Elsevier's Grand Challenge contest three years ago. We conducted a formal usability study of Panorama in 2013, and then enhanced the software to include bar charts and the ability to run on the Mac OS X operating system.

We extended Panorama to provide Annotation Concepts. Clicking on text in an interactive publication prompts an NLM servlet (RIDeM, developed in-house in 2004) to identify the corresponding Unified Medical Language System, or UMLS, concepts. The servlet returns an XML (eXtensible Markup Language) file to Panorama, which parses it to provide the preferred UMLS term and semantic group, and it provides linkouts for MedlinePlus, eMedline, Family Doctor, and other resources. We are continuing to develop additional features to group concepts by semantic relationships and other factors.

To avoid the need for the large downloads required by our original (desktop) approach, we investigated several Web-based methods for reading interactive publications. In 2012 we developed our first browser-based version (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. In 2013, we improved and updated Panorama Lite features. Besides offering easy and intuitive usage, this version has better line-chart and graph support and includes tables and subsets similar to the original Panorama. Panorama Lite also features a unique map view that can present data at the county, state, and country levels in a color-coded form so users can visualize geographic patterns relatively easily.

Over the past two years, we've taken this project on the road, collaborating with two organizations to create interactive publications from their traditional, static ones: a publisher (ProQuest) and a government agency (the Centers for Disease Control and Prevention's National Center for Health Statistics, or CDC/NCHS). We created two interactive papers for ProQuest from one of their open-source journals (*Sustainability: Science, Practice and Policy*), and the company announced the launch of these papers for public use in a press release.

For CDC/NCHS, we converted two issues of their key document — the 2011 and 2012 *Health US In Brief* reports — to interactive form and hosted them on our Web site. *In Brief* contains summary information on the health of the American people, including mortality and life expectancy, morbidity and risk factors such as cigarette smoking and overweight and obesity, access to and use of health care, health insurance coverage, supply of healthcare resources, and health expenditures. We also converted another important CDC/NCHS document, the Data Brief No. 115 March 2013, *Death in the United States, 2011*. We expanded Panorama's functionality to support drill-down pie charts for all these interactive papers, and we added usability support for standard deviation charts for *In Brief* using HTML5/javascript.

Screening Chest X-rays for Tuberculosis in Rural Africa

In FY2013, we continued our collaborative project with AMPATH (Academic Model Providing Access to Healthcare), an organization supported by the U.S. Agency for International Development (USAID) that runs the largest AIDS treatment program in the world. Through this project, we're conducting imaging research and developing systems to support NIH efforts to improve global health. Our objective is to use our 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) and other lung infections prevalent in patients with HIV. We provided AMPATH with lightweight digital X-ray units that are easy to transport in rural areas. The AMPATH staff will take chest X-rays (CXR) of people and screen them for the presence of disease.

In the past year, one of our X-ray units was mounted in the MOI University Hospital in the town of Eldoret in western Kenya. The images from the X-ray units are in a standardized DICOM radiological-image format. Through advances in technologies for Web-based Access to DICOM Objects (WADO) and the implementation of long-range WiFi in western Kenya, images acquired in the field can be stored in the PACS (Picture Archiving and Communications System, a database system used in hospitals to store medical images) housed at the AMPATH building on the hospital grounds in Eldoret.

Because of the lack of sufficient radiological services in western Kenya, we've been focusing our in-house research effort on developing software that automatically screens the CXR images for disease. Our researchers are developing machine-learning algorithms to automatically segment the lungs; detect and remove ribs, heart, aorta, and other structures from the images; and then detect texture features characteristic of abnormalities, which allows us to discriminate abnormal from normal cases. These machine-learning algorithms, which allow computers to learn so they can do a task without being programmed to do it, require large sets of example X-rays. After receiving an internal review board (IRB) exemption, we explored many options for acquiring chest X-ray (CXR) training sets. We now have about 400 CXRs from Montgomery County's TB Control Program, 850 from a source in India, 2,000 from a hospital in China, 8,200 from Indiana University, and 250 from an open-source set from Japan.

The usable number of CXRs in our collection is less than the total. This is partly because of the marginal

quality of many images and the inclusion of lateral views, which we're not considering yet. To create a training set, we manually validate and annotate the images, and in FY2013, we completed this process for the images from Indiana University. To ensure the privacy of the X-rayed subject, we discarded images that included any form of personally identifiable information, such as cardiac pacemakers with inscribed serial numbers, or dental or tubes that are common in hospital-based radiological imaging, since they tend to confuse the classifiers.

Using these X-rays for training and testing, we developed algorithms for detecting relevant anatomy in the chest regions such as lungs, ribs, and the heart. To ensure the acquisition of high-quality images, we are using rib structures to develop other algorithms that detect the planar and positional rotation of the patient because such rotation can lead to incorrect diagnosis. Being able to detect the heart not only allows us to separate it from the lung regions, but we can then also detect abnormalities such as cardiomegaly (enlarged heart) that can be a precursor to congestive heart failure.

In 2013, we significantly improved the lung-segmentation algorithm with a novel atlas-based method that is 96 percent accurate. The algorithm is now better able to detect lung boundaries in CXR images for lungs exhibiting disease pathology, and by increasing the algorithm's computational efficiency we've also enabled it to run much faster (30 seconds versus several minutes). Radiologists from the NIH Clinical Center, Yale University, and the University of Missouri in Columbia helped us by annotating pathology in some of our images. We used these annotated images to train our SVM-based classifier, which uses several features extracted from the X-rays as input, such as histograms of intensity, gradient magnitude and orientation, shape, and curvature. On the basis of these input features, the SVM returns a confidence value, allowing an operator to inspect cases in which the classifier is uncertain. We also compared the performance of the algorithm classifier with that of human experts. We found that they perform similarly (87 percent accurate), but the classifier tends to be more sensitive, yielding nearly twice as many false positives. While not ideal, this oversensitivity does prevent overlooking X-rays that show disease, and it's useful in a resource-constrained setting. We are continuing to research methods for advancing classifier performance by sampling image patches.

Computational Photography Project for Pill Identification (C3PI)

Launched in September 2010, the Computational Photography Project for Pill Identification (C3PI) is a start on an authoritative, comprehensive, public digital-image inventory of the nation's commercial prescription, solid-dose medications. Working with expert consultants, we're creating a collection of photographs of prescription tablets and capsules, confirming that the images match the description of the medication, and developing and matching the images of the samples to relevant metadata (such as size descriptions, dimensions, color, and the provenance of the sample).

In FY2013, we increased our online collection to more than 70,000 images of nearly 2,000 pill (solid-dose) samples pharmaceuticals from more than 150 manufacturers and distributors. The team generates high-resolution, high-quality pill images from a variety of lighting conditions. The long-term goal of our collaboration is to facilitate the development of computerized tools to help identify medication based on features such as shape and color.

Also in FY2013, as part of a public launch of the C3PI project, we worked with the contractor to create two public online repositories: http://rximage.nlm.nih.gov (which links to RxNav) and http://splimage.nlm.nih.gov. Through these Web sites, we can distribute images of oral, solid-dose medications to the public and to pharmaceutical manufacturers, respectively.

As part of cost management and to reflect the current needs of the C3PI project, we established a new contract, conducted a fair and open competition, and awarded the contract to Medicos Consultants at the end of FY2013. In FY2013, we completed all the research to establish the imaging, database, delivery, and storage protocols for structured product label (SPL) IMAGE files, and entered production mode for C3PI. This was a joint development effort with contributions by both NLM staff and the contractor. In FY2014, we'll acquire data using these protocols.

The Visible Human Project

The Visible Human Project image datasets serve as a common reference for the study of human anatomy, a set of common public domain data for testing medical imaging algorithms, and a test bed and model for constructing image libraries that can be accessed through networks. These datasets are available through a free license agreement with the NLM. We distribute them in their original or in PNG (*Portable Network Graphics*, a raster graphics file format that supports lossless data compression**)** format to licensees over the Internet at no cost and on DVDs for a duplication fee. Almost 3,450 licensees in 64 countries are applying the datasets to a wide range of educational, diagnostic, treatment-planning, virtual-reality, and virtual-surgery uses, in addition to artistic, mathematical, legal, and industrial uses. In FY2013, we continued to maintain two databases for information about how people are using the Visible Human Project: one for information about the license holders and their intended use of the images, and the other for information about the products the licensees provide NLM as part of the Visible Human Dataset License Agreement. More than 1,000 newspaper, magazine, and

radio pieces have featured the Visible Human Project since we released the first dataset in 1994.

3D Informatics

We continue to address problems encountered in the world of three-dimensional and higher-dimentional, time-varying imaging through our 3D Informatics Program (3DI). We provide ongoing support for image databases, including the National Online Volumetric Archive (NOVA), which NLM created in 2003 and continues to host. NOVA contains 3D data from numerous medical institutions, including the Mayo Clinic Biomedical Imaging Resource and the Walter Reed Army Medical Center Radiology Department. In FY2013, 3DI added data to NOVA from an ion-abrasion scanning electron microscope from our collaboration with NCI in high-resolution electron microscopy. We continue to serve a broad community with these data and have become leaders in the distribution of medical data to the public.

Throughout FY2013, we continued our collaboration with NCI's Laboratory for Cell Biology and with teams within LHNCBC to visualize and analyze complex 3D volume data generated through dual-beam (ion-abrasion electron microscopy) and cryo-electron tomography. This work combines high-performance computing with life sciences research, related to the detection and prevention of cancer and infectious diseases. The resulting visuals have provided insights about the character of several immunological cells, cell structures, and the cells' interaction with pathological viruses, including HIV.

We continued our commitment to processing data collected through transmission electron tomography. We successfully published the results of software-development research that uses GPUs for high-performance computing for sub-volume averaging and reconstruction. (The image data collected are equivalent to the serial sections of a CT scan. These sections can be put back together to "reconstruct" the original "volume." Any mathematics performed on part of the "volume" is called "sub-volume.") We're working now to develop emerging methods into mature software that can extend this work on sub-volume reconstruction to single-particle microscopy. If successful, the software will expand our impact to a wider range of molecular targets in systems biology.

This past year, we also helped supervise segmentation efforts for data from ion-abrasion electron microscopy in a study of the disposition of malaria pathogens (that is, where they wind up) in normal human blood cells. We're evaluating the impact of heterogeneous sickle cell hemoglobin on the infiltration, geometry, and spatial relationships of the pathogens in affected erythrocytes. Other work in related areas includes the segmentation and study of vaccinia viruses in normal and mutant human liver cells.

Insight Tool Kit

The Insight Toolkit (ITK) is a public, open-source algorithm library for segmenting and registering high-dimensional biomedical-image data. The current official software release is ITKv4.2.2. More than 845,000 lines of openly available source code comprise ITK, and with it, people can access a variety of image-processing algorithms for computing segmentation and for registering high-dimensional medical data. ITK runs on Windows, Macintosh, and Linux platforms, so it can reach across a broad scientific community. It is used by more than 1,500 active subscribers from 40 countries. A consortium of university and commercial groups, including our intramural research staff, provides support, development, and maintenance of the software.

ITK remains an essential part of the software infrastructure of many projects across and beyond 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 applied programming interface (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 Mayo Clinic's Analyze and SCIRun from the University of Utah's Scientific Computing and Imaging Institute, and through developing a new release of VolView, free software for medical volume image viewing and analysis.

Our ongoing efforts in 2013 included expanding and continuing to develop Simple ITK, or sITK. Developed primarily by programmers at NLM, sITK is a simplified layer built on top of ITK. It's intended to facilitate ITK's use in rapid software development and education through the support of scripting languages, primarily Python. The sITK interface conceals the structural and design complexities of ITK, enabling more straightforward, procedure-based programming. Designed to be an interpreted scripting system, sITK supports a typeless, polymorphic data model, thus simplifying the use and expression of ITK in image-analysis education. The current software release of Simple ITK is version 0.7.

3D Printing

In FY2013, the 3DI group participated in the HHS Ignites challenge by proposing, with our colleagues at the National Institute of Allergy and Infectious Diseases (NIAID) and the National Institute of Child Health and Human Development (NICHD), an online repository called the *NIH 3D Print Exchange*. Designed to supply biomedical shape files for education and research to a growing audience of users worldwide, the exchange provides public software built from NLM's Insight and the Visualization Toolkits to automatically generate printable models from X-ray CT data. The project Web site is currently online in beta-testing mode, and a national advisory board is assisting with oversight, testing, and guidance for the project.

The 3DI group also continued to investigate the use of rapid-prototyping technologies in radiology with partners at NIAID. 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. A new set of models is under development, including dosimetry models from CT scans of small animals.

Open-*i*: *Image and Text Indexing for Clinical Decision Support and Education*

A picture is worth a thousand words, especially in medical research and clinical practice! Most people can understand complex biomedical concepts more easily if they are presented visually: through radiographic images, photographs of organs, sketches, graphs, or charts. This idea motivates our project, part of the LHNCBC Clinical Information Systems effort, which exploits ongoing research in both natural language processing and content-based image retrieval by processing both text and image features. We developed the Open-*i* system for finding images and figures in published literature or other sources. Open-*i* system enables users to search and retrieve medical citations from 450,000 open-access articles in PubMed Central® (PMC). Users may search by text queries as well as by example images. They can filter images by type (e.g. X-ray, graph, etc.), filter journals by clinical specialty, and rank papers by clinical task (e.g., treatment).

Open-*i* was released to the public in 2012 and is the first production-quality system of its kind in the biomedical domain, Open-*i* gives medical professionals and the public access to images contained in biomedical articles that are highly relevant to their query, as well as a summary of the articles' take-away messages. The system is available 24/7, and it can handle about 20,000 interactions per day in real time. The quality of the information delivered by Open-*i* has been evaluated in international medical-image-retrieval "Cross Language Evaluation Forum" competitions (ImageCLEF), in which the system consistently ranks among the best. For example, the system demonstrated the best retrieval results in a 2013 ImageCLEF that attracted participants from academia, industry, and clinical settings.

In 2013 we added a collection of clinical documents to the Open-*i* open-access biomedical articles. This collection, provided by Indiana University, consists of chest X-ray examinations and accompanying de-identified reports for about 4,000 patients. To prepare the collection for indexing and public release, the Open-*i* team has reviewed the university's de-identification of the DICOM images and radiology reports and removed images with any possible risk of re-identification (for example, images of teeth).Then, we converted the remaining images to JPEG format for indexing. We also indexed the text reports with NLM's MeSH® (medical subject heading) terms, and converted the reports to Open-*i*-enriched citations. The resulting collection is indexed and searchable in Open-*i* either by itself or in combination with the scientific publications.

This past year we also increased the collection of images in Open-*i* to 1.3 million from PubMed Central articles (from 450,000 in 2012) and 8,000 from the 4,000 Indiana University radiology reports. The demand for Open-*i* services nearly doubled in 2013 — about 20,000 total visitors, bots, and auxiliary pages a day, representing (compared with 11,000 in 2012) — to accommodate this increase, we upgraded the system architecture, refactored the XML structure, and redesigned the indexes. In addition to online searches, Open-*i* provides batch-retrieval services for researchers who need access to images. For example, the University of Pittsburgh uses the services to add illustrations to their oral squamous cell carcinoma dataset.

In response to users' requests in 2013, Open-*i* introduced a search-history list that allows users to see their previous searches. Adjustments to our search algorithm led to an improvement in modality classification from 85 percent to 90 percent. We also modified the auto-complete suggestions in searches, based on information about the presence and frequency of the UMLS terms in the collection of documents indexed in Open-*i*.

The evolution of Open-*i* continues to be supported by research in several key areas previously described in this chapter.

In other work this past year, we took steps toward building a visual structural framework for organizing information called a visual ontology. We 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 (TTP)

The goal of this project is to give people access to historically significant and previously inaccessible books in medicine and the life sciences. We build 3D models for books and develop animation techniques that let users touch and turn page images in a realistic way on touch-

sensitive monitors in kiosks at NLM or tablets using a high-resolution ("Retina") iPad app, or click and turn if they're using an online version. We've also built a different 3D model for a "scroll"-type document and applied it to the 1700 BC Edwin Smith medical papyrus, which can be "touched" (or clicked) and "rolled out." The TTP Web site is very popular, attracting about 300,000 page views a month. The iPad app is also popular: in 2013, 3,270 people downloaded the app for the first time (new users), and more than 16,751 people who already had the app updated it with the newer version of the software, which indicates a steady user base.

The Turning the Pages kiosks at NLM and the NLM Web site now present 11 rare books. In 2013, we added four books to the iPad version, bringing that total to 10, with only the Smith papyrus yet to be added. Now included in the TTP Web and iPad versions is Elizabeth Blackwell's *A Curious Herbal* (published between 1777 and 1779), previously available only at the NLM kiosks. We had to recreate all the graphics and update the animation software for that project. We also added new features to the iPad version in 2013, such as bookmarks and contextual zoom for curators' notes. To reduce project costs, we modified our production pipeline to capture images in-house rather than using commercial scanning companies. Right now, we're creating the TTP version of a Mongolian Prayer Scroll. We're also investigating options to create a TTP version of the now-historic genetic charts, created by Nobel Laureate Marshall Nirenberg, that elucidate the process of translating the four-letter code of DNA into the 20-letter alphabet of amino acids, the building blocks of proteins.

For the longer term, we're studying ways to develop a reactive 3D implementation system for TTP and investigating tools for this purpose, such as Unity, Coco's 3D, and the Unreal Engine. The advantages of a real-time 3D system are:

- We can produce 3D versions of each book more quickly,
- Other institutions can use our software to create their own interactive books, and
- We can discover new functionalities, such as rotating a book 360° and turning multiple pages at once.

In addition, anticipating the next generation of kiosk design, we're investigating newer display technologies, such as multi-touch monitors, which enables users to use two or more fingers to perform tasks such as zooming in and out.

Natural Language Processing and Text Mining

Medical Article Records System (MARS)

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

While the bulk of citations now come to NLM directly from publishers (in XML format), nearly 820 journal titles are provided only in paper form. These papers are processed by the Medical Article Records System, or MARS, which we launched in 1996. MARS combines document-scanning, optical character recognition (OCR), and rule-based and machine-learning algorithms to extract citation data from paper copies of medical journals for MEDLINE. Our algorithms extract the data through a pipeline process: segmenting page images into zones, assigning content labels to the zones (title, author names, abstract, grant numbers, etc.), and pattern matching to identify these entities.

We manage and continually improve the MARS system. For example, we're introducing three new features: (1) an expanded MEDLINE character set, (2) capability for the Edit operators to correct errors made by the automated zoning process, and (3) a new user-interface design for large-screen monitors. We have completed the software implementation and the integration test for the first two features, which are now being deployed to the production system used by NLM Library Operations.

Citations that come to NLM in electronic form from publishers often contain errors or have missing content. Missing items include databank accession (identifying) numbers from databases such as GenBank, NIH grant numbers, grant support categories, investigator names, and information about the links between articles and the comments submitted in response to them. The capture of investigator names can be especially difficult because some articles contain hundreds of such names, and capturing the articles that "comment on" another paper (usually an editorial or a review article) requires operators to open and read other articles related to the one being processed. To automatically extract these fields from online articles, we developed the Publisher Data Review System (PDRS), whose subsystems are based largely on machine-learning algorithms such as the support vector machine, or SVM.

PDRS went into production in early FY2012 for open-access articles in NLM's PubMed Central. To extend automated data extraction to *all* online journals on publishers' sites, including the ones with restrictive copyrights, we're developing IMPPOA (In-Memory Processing for Publisher Online Articles). This is a system based on the PDRS platform and its machine-learning algorithms, but implemented to process articles in RAM memory, which is temporary. IMPOAA:

- Provides data missing from the XML citations sent in directly by publishers.
- Corrects errors in publishers' data by extracting data from the articles on their sites and comparing these with the data sent to NLM.
- Extracts data from articles for which publishers do not send in citations at all.

Because this new system avoids downloading the articles to a disk drive, we expect IMPOAA to eliminate publishers' concerns about copying articles into an external system disk.

The systems outlined above rely on underlying research in image processing and lexical analysis that also enables the creation of new initiatives for applying these techniques, such as the ACORN project (also known as the Automatically Creating OLDMEDLINE Records for NLM project), described below.

Digital Preservation Research

The long-term preservation of documents in electronic form is a critical task for NLM as well as other libraries and archives. The goal of this LHNCBC project is to investigate and implement techniques for key preservation functions, including automatically extracting metadata to enable future access to the documents, storing the documents and metadata, and discovering knowledge in the archived material. To provide a platform for this research, we built and deployed a System for Preservation of Electronic Resources (SPER). SPER builds on open-source systems and standards (e.g., DSpace, RDF), and incorporates in-house-developed modules that implement key preservation functions.

Our research focuses on two collections. One is a historic medico-legal collection of 67,000 early twentieth-century court documents from the FDA. These are "notices of judgment" issued by courts against companies that were indicted for misbranding or adulterating foods, drugs, or cosmetics. They offer insights into legal and governmental history dating from the 1906 Food and Drug Act, and illustrate regulatory impacts on public health. Curators in the NLM History of Medicine Division are using SPER to preserve the FDA documents, and in 2013 they processed more than 15,000 of them. These documents and their metadata are in a publicly accessible NLM Web site.

The second collection, from NIAID, is a set of conference proceedings of the US-Japan Cooperative Medical Science Program (CMSP) Cholera and Other Bacterial Enteric Infections Panel, an international program conducted over a 50-year period from 1960 to 2011. For this collection, our activities include: (1) building a full repository for 2,800 research articles on cholera and 8,000 references on CMSP participants such as authors, panelists, attendees, and study section reviewers, followed by (2) developing a portal where the public can go to search for research articles, institutes, and authors. To support these activities, we developed techniques for automatically extracting three different types of metadata from the CMSP documents:

- Publication metadata with titles, authors, and their affiliated institutions from research articles.
- Investigator metadata with name, role, designation, and affiliation of each person from the conference proceedings rosters.
- Study section metadata with names and affiliations of CMSP program reviewers from separate study section rosters.

We then used the metadata to implement data-analysis functions for discovering patterns and trends in factors such as important drugs, discoveries, investigators, and international collaborations under the CMSP program over its 50-year span.

In FY2013, we explored building a knowledgebase from the metadata of an archived collection and performing semantic queries against such archives. We hoped to discover important domain-specific information, thus generalizing the data-analysis capability. We selected open-source tools to create a model based on the OWL/RDF structural framework for organizing information for a given collection characterized by its metadata fields and their interrelations, and we developed new tools to generate a knowledgebase from the stored metadata using that framework. We used techniques to semantically query such knowledgebases through a Web browser and then to graphically display the results.

We applied this methodology to the CMSP collection by building a complete CMSP knowledgebase and performing semantic queries to obtain various patterns and trends of interest. We also developed an ontology model for the FDA collection that can be used to generate the corresponding knowledgebase once the collection is fully archived.

Automatically Creating OLDMEDLINE Records for NLM (ACORN)

NLM would like to expand MEDLINE to include all bibliographic records beginning in 1879, when *Index Medicus* was first developed. The earliest citations exist only in printed paper form, and NLM Library Operations (LO) has collected many of these with considerable manual effort. To translate these paper indexes into electronic format, we designed the ACORN system, which combines scanning, image enhancement, optical character recognition (OCR), image analysis, pattern matching, and related techniques to extract electronic versions of the printed indexes. The fact that the printed versions used old typefaces, fonts, and a mix of different languages yields inaccurate OCR results. To overcome this problem in one of the indexes (Quarterly Cumulative Index Medicus, or QCIM), we developed a novel pattern-matching technique that automatically finds and compares two versions of every citation from the subject and author listings, thereby minimizing the OCR errors encountered in each version. In addition, our system searches MEDLINE to avoid duplicating records that already exist.

ACORN has three main components: Scanning and Quality Control, Processing, and Reconciliation (that is, when operators verify the extracted information). We completed and delivered the first component to NLM Library Operations on April 2013, and by the end of 2013 the LO staff had scanned and completed quality-control

checks of 16 of the 60 QCIM volumes. We are still developing the other two components and will release them for production in 2014.

Indexing Initiative

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

NLM Library Operations (LO) MEDLINE citations indexers regularly (and increasingly) use the MTI system. To facilitate their indexing, MTI provided recommendations for 673,428 articles in FY2013 as an additional resource available through the Data Creation and Maintenance System (DCMS). Because of the recent addition of subheading attachment recommendations, indexers now have the option of accepting MTI heading-subheading pairs in addition to unadorned headings. Our developers also created specialized versions of MTI to assist in the indexing of the NLM History of Medicine book collection and for general cataloging. Due to its success with certain journals, MTI was designated as the first-line indexer (MTIFL) for those journals. As a "first-line" indexer, MTI indexing is also still subject to human manual review. The number of MTIFL journals will grow gradually and should prove to be a time and money saver for NLM.

In FY2013, we added 75 new journals (for a total of 120) to the MTIFL program, which included 9,771 articles. We are collaborating with LO to evaluate how well MTI is performing on indexing the journals that are already part of the MTIFL program by computing standard information-retrieval measures (recall, precision, and f-measure) and then comparing MTI's indexing recommendations with the final, official MEDLINE indexing. We also work with LO to identify future MTIFL journal candidates.

In FY2013, MTI provided the primary baseline for the international BioASQ challenge for biomedical semantic indexing and question answering (http://www.bioasq.org/). The aim of the challenge is to make biomedical text more accessible to researchers and clinicians. The MTI indexing results provided one of the baselines used in the "large-scale online biomedical semantic indexing" part of the challenge, which is designed to parallel the human indexing currently being done at NLM. The II team provided help and guidance in developing the list of journals used in the challenge, as well as the baseline results. MTI will also provide a baseline for the second year of the BioASQ challenge in 2014.

The II team collaborated with LO to develop a new system to help with indexing LO's NLM Technical Bulletins. The system automatically supplies keywords for all Technical Bulletins, which are then manually reviewed before publication. In FY2013, our new system processed 1,361 Technical Bulletins, including all archived Technical Bulletins from 1997 to the present. A comment from an LO staffer illustrates how effective the new system has been for them: "The team is using the system to index all Technical Bulletin articles now, streamlining our workflow and making our assigned terms consistent. We are saving a lot of time with each article!"

MetaMap is a system for identifying concepts within text documents. It is a critical component of the MTI system and is also used worldwide in bioinformatics research. In 2013, we significantly improved its processing speed, added XML output, implemented a way to determine whether a statement is positive or negative, e.g., "He has cancer." vs. "He does not have cancer" (known as negation identification), and enabled users to supply their own acronyms-abbreviations list. In 2013, we migrated MetaMap's legacy lexicon module to a Java-based implementation. MetaMap is available on Windows, Macintosh and Linux platforms. Users can build their own datasets with the MetaMap Data File Builder and run their local version of MetaMap to process documents containing sensitive data via either an embedded Java API (applied programming interface) or a UIMA (unstructured information management architecture) wrapper. In FY2013, users downloaded about 2,000 copies of MetaMap and 900 copies of the Java API and UIMA wrapper.

RIDeM/InfoBot

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

The related InfoBot project enables a clinical institution to automatically augment a patient's electronic medical record (EMR) with pertinent information from NLM and other information resources. The RIDeM API developed for InfoBot allows just-in-time access to patient-specific information to be integrated into an existing EMR system. Such patient-specific information includes medications linked to lists of medications for each patient, or formularies, and images of pills, evidence-based search results for patients' complaints and symptoms, and

MedlinePlus information for patient education. For clinical settings without access to the API, a Web-based interface allows information requests to be entered manually.

The InfoBot API integrated with the NIH Clinical Center's EMR system (CRIS) has been in daily use through the *Evidence-Based Practice* tab in CRIS since July 2009. In 2013, the tab was accessed 609 times a month, on average, by more than 1,380 unique users at the NIH Clinical Center.

Consumer Health Information and Question Answering (CHIQA) system

NLM's customer services receive about 90,000 requests a year. In 2012, we started to investigate the possibility of automating the process of answering these consumer health questions. In 2013, we developed and evaluated a prototype Consumer Health Information and Question Answering system (CHIQA). The prototype can classify the incoming requests as either questions about health problems or requests to correct MEDLINE citations. Once the request type is recognized, CHIQA generates an answer and submits it to NLM's reference staff for review. For MEDLINE correction requests, the system automatically finds and retrieves the citation that set off the request, extracts relevant information, and generates an answer. The prototype also understands simple frequently asked questions about causes, treatments, and prognoses of diseases. For these questions, CHIQA finds relevant articles from NLM consumer resources, such as Genetics Home Reference and MedlinePlus, and uses sections of the articles to answer the questions.

De-identification Tools

De-identification allows people to conduct research on clinical narrative reports. We are designing a Clinical Text De-identification (CTD) system that will remove protected health identifiers from narrative clinical reports. The provisions of the Privacy Rule of the Health Insurance and Accountability Act require the removal of 18 individually identifiable health information elements that could be used to identify the individual or the individual's relatives, employers, or household members.

We completed a version of the software system to be tested at the NIH Clinical Center. This version de-identifies clinical narrative text in a form of electronic messaging known as Health Level Seven, Inc. (HL7) version 2. It can use information embedded in various HL7 fields as well as externally provided information, such as list of names of the healthcare providers at NIH.

We're using the Visual Tagging Tool (VTT) we designed to produce gold standards against which to test the CTD system. Although we designed the VTT specifically to help the CTD identify protected health information (PHI), the natural language processing (NLP) community is already also using it for other types of lexical tagging and text annotation.

By the end of 2013, we amassed a collection of 21,849 clinical reports of 7,571 patients – in which a human reviewer manually labeled every piece of individually identifiable health information. We can use this collection as the gold standard for testing the CTD system.

Librarian Infobutton Tailoring Environment (LITE)

Infobuttons (http://www.infobuttons.org) are links from one information system to another that anticipate users' information needs, take them to appropriate resources, and help them retrieve relevant information. They are mostly found in clinical information systems (such as electronic health records (EHRs) and PHRs) to give clinicians and patients access to literature and other resources that are relevant to the clinical data they are viewing. The NIH Clinical Center Laboratory for Clinical Informatics Development has worked with Health Level Seven, Inc. (HL7) — an electronic messaging standards development organization — to develop an international standard to support the communication between clinical systems and knowledge resources. MedlinePlus Connect currently provides an HL7-compliant query capability.

To increase the usefulness of infobuttons, they are typically linked not to a specific resource, but instead to an "infobutton manager" that uses contextual information (such as the age and gender of the patient, the role of the user, and the clinical data being reviewed) to select the most applicable resources from a large library of known resources. The infobutton manager customizes the links to those resources using appropriate data from the context, and then presents the user with a list of those links. The NIH Clinical Center Laboratory for Informatics Development is working with investigators at the University of Utah and the Department of Veterans Affairs to establish a freely available, HL7-compliant infobutton manager, known as "Open Infobutton" (http://www.openinfobutton.org), that can be a national resource for electronic health record developers and users. With the Open Infobutton, clinicians and patients will be able to obtain the health-related information they need, when and where they need it.

Infobutton managers, including Open Infobutton, require knowledgebases to do their customization work. The knowledgebases are very institution-specific. We developed the Librarian Infobutton Tailoring Environment (LITE), a user-friendly tool that can be used by an institution's medical librarians and provides Open Infobutton with the knowledge it needs to customize its responses to requests from that institution. The system is in beta testing at the University of Utah (http://lite.bmi.utah.edu). In 2013, we transferred the maintenance of LITE to the University of Utah, which will continue to make it open-access and will also develop the software to make it part of an open-source package that can be installed at any institution. A user-evaluation

project is now under way in collaboration with Ohio University.

Terminology Research and Services

The Patient Data Management Project (PDM) brings together several activities centered on lexical issues, including developing and maintaining the SPECIALIST lexicon and lexical research. Those lexicon and lexical tools support key NLM applications. A package of lexical-tool applications underlies the MetaMap algorithm we use to find UMLS concepts in biomedical text and to automatically index MEDLINE abstracts. We distribute the lexicon and lexical tools to the medical informatics community as free open-source tools.

In FY2013, we provided support to 35 internal users, 14 U.S. domestic users, and eight international users. We also enhanced the derivational-variants function of the lexical tools. Derivational variants are words related by a word-formation process like suffixation, prefixation, or conversion (change of category). The enhanced derivational-variant system will be part of the UMLS 2014 release. The 2014 release of the SPECIALIST lexicon contains 476,856 records representing more than 875,000 forms, an increase of 6,865 records from the 2013 release. Many of the new records are derived from de-identified clinical records from our own de-identification project and from our work with the MIMIC-II database.

Medical Ontology Research

The Medical Ontology Research (MOR) project focuses on basic research on biomedical terminologies and ontologies and their applications to natural language processing, clinical decision support, translational medicine, data integration, and EMR interoperability. During FY2013, we investigated the use of terminological resources for mapping between rare-disease information sources, and we developed a framework for aligning pharmacologic classes between MeSH and Anatomical Therapeutic Chemical (ATC) classification. We supported the NLM Value Set Authority Center (VSAC) – a crucial part of the quality measures stemming from the meaningful use regulations – by proposing metrics for assessing the quality of value sets in clinical quality measures. Finally, in collaboration with the Center for Drug Evaluation and Research at the FDA, we developed methods for extracting information about adverse drug events from MEDLINE indexing.

Research activities this year resulted in one journal article, five papers in conference proceedings, seven editorials, abstracts, and posters, one book chapter, and five invited presentations. We continue to collaborate with leading ontology and terminology centers, including the National Center for Biomedical Ontology, the International Health Terminology Standards Development Organization (SNOMED CT), and the World Health Organization (ICD-11, the eleventh version of the International Classification of Diseases).

Semantic Knowledge Representation

The Semantic Knowledge Representation (SKR) project conducts basic research, based on the UMLS knowledge sources, in symbolic natural language processing. A core resource is the SemRep program, which extracts semantic predications (relationships — such as interacts with, treats, causes, inhibits, and stimulates — between drug and disease, gene and gene, gene and disease, drug and, etc.) from biomedical text. SemRep was originally developed for biomedical research. We're developing a way to extend its domain to influenza epidemic preparedness, health promotion, and health effects of climate change. In FY2013, we made a downloadable version of SemRep available to the public.

SemRep finds biomedical-related semantic relationships in MEDLINE, and then our Semantic MEDLINE Web application manipulates those relationships. The SKR project maintains a database of 60 million SemRep predications extracted from all MEDLINE citations that is available to the research community. This database supports Semantic MEDLINE, which integrates PubMed searching, SemRep predications, automatic summarization, and data visualization. The application helps users manage the results of PubMed searches by creating 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 the UMLS Metathesaurus). The Semantic MEDLINE technology was recently adapted for analyzing NIH grant applications, allowing NIH portfolio analysts to track emerging biomedical research trends and identify innovative research opportunities.

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

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

ClinicalTrials.gov

Established in 2000, ClinicalTrials.gov makes public comprehensive information about registered clinical research studies. It receives more than 95 million page views and hosts about 980,000 unique visitors per month. Nearly 13,000 study sponsors, including the federal government, pharmaceutical and device companies, and academic and international organizations, submit data to ClinicalTrials.gov through a Web-based Protocol Registration System (PRS). At the end of FY2013, the site had nearly 154,000 research studies, conducted in all 50 states and in more than 180 countries. Approximately one-third of the studies are still open to recruitment. For the remaining two-thirds, the recruitment phase is over or the study has been completed. More than 10,400 of the closed

studies display summary-results tables describing primary and secondary outcomes, adverse events, and characteristics of the participants studied.

In FY2013, new registrations of clinical trials were submitted at an average rate of 400 records per week, an increase of 8 percent from FY2012. The average rate of new results submissions was about 70 per week, consistent with FY2012. We can attribute the continued growth in the use of ClinicalTrials.gov to US laws that require registering and reporting the results of clinical trials, as well as international recognition of the scientific and ethical importance of registration and reporting results. The combined registry-and-results database provides information about ongoing and completed clinical research for patients, healthcare providers, and policy decision makers.

In FY2013, ClinicalTrials.gov staff continued to implement and educate the public about the most recent clinical trial law, Section 801 of the Food and Drug Administration Amendments Act of 2007 (FDAAA 801). We have been working with the NIH Office of the Director, other NIH Institutes and Centers, and the FDA on a Notice of Proposed Rulemaking (NPRM) that will elucidate the requirements of FDAAA 801 and solicit public comment on key implementation issues.

Following the launch of the redesigned public Web site in FY2012, we started evaluating the public's experience with the redesign and shifted our focus to evaluating and enhancing the ClinicalTrials.gov data-entry site, the PRS. Sponsors and investigators use the site to submit, update, and maintain information about their studies. On the basis of comments from users to an online survey and other user feedback, we implemented user-interface improvements intended to streamline the data-entry process. We started a usability study of the PRS in FY2013, and we'll use its findings for further enhancements in FY2014. We continued providing targeted education and outreach on the results database and submission requirements through hosting three on-site training workshops, presenting at conferences, participating in working groups, and publishing in journals. ClinicalTrials.gov research projects in FY2013 included:

- Characterizing prematurely terminated trials registered at ClinicalTrials.gov.
- Illuminating reasons for early termination.
- Finding out how many and what types of summary results are available from such trials in the ClinicalTrials.gov results database and the published literature.

In April 2013, the LHNCBC Board of Scientific Counselors reviewed the ClinicalTrials.gov program and praised it. The program continued to provide technical advice and collaborate with other clinical study registries, professional organizations, funders, and regulators on working toward developing global standards for trial registration and for reporting to results databases. For example, a key activity continues to be working with the European Medicines Agency (EMA) on developing a common set of data elements for results submission to both ClinicalTrials.gov and the EMA results database, which is being developed for release in FY2014.

Genetics Home Reference (GHR)

The Genetics Home Reference (GHR) Web site offers high-quality information about genetic conditions and the genes and chromosomes related to those conditions. It answers the public's questions about human genetics and the rich technical data from the Human Genome Project and other genomic research. At the end of FY2013, the GHR included user-friendly summaries of more than 2,100 genetics topics, including more than 900 genetic conditions; about 1,250 genes and gene families; all the human chromosomes; and mitochondrial DNA. GHR 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.

GHR celebrated its 10th anniversary in FY2013. In the past year, we expanded the project's genetics content for consumers, adding 242 new condition, gene, and gene-family summaries and new Help Me Understand Genetics pages about genetic susceptibility and informed consent. In FY2013, the site averaged almost 42,500 visitors per day and about 34.9 million hits per month (an increase of 52 percent and 26 percent, respectively, from FY2012).

In FY2013, we integrated GHR results into NLM's MedlinePlus Connect. This service enables electronic medical records and other applications that use MedlinePlus Connect to retrieve GHR summaries (along with MedlinePlus content) by using code queries from SNOMED CT (the Systematized Nomenclature of Medicine Clinical Terms). GHR topics, each of which can map to multiple SNOMED CT codes, currently map to more than 2,300 SNOMED CT codes.

GHR continued its formal collaboration with Genetic Alliance, an umbrella organization for condition-specific genetics interest groups, to update existing GHR Web site content and to track new clinical and research developments about particular genetic conditions. We updated about 70 existing GHR topics through this initiative in FY2013. We also performed outreach activities last year to increase public awareness about GHR. We presented the Web site to several groups, including health and science journalists who visited NLM as part of the Association of Health Care Journalists–NLM Fellowship program, clinical and molecular fellows at the National Human Genome Research Institute (NHGRI), and conference attendees at the annual meetings of the National Society of Genetic Counselors and the American Medical Writers Association.

Profiles in Science Digital Library

The *Profiles in Science*® Web site showcases digital reproductions of items selected from:

- The personal manuscript collections of 33 prominent biomedical researchers, doctors, public health practitioners, philanthropists, political leaders, and other people who provided resources, removed barriers, and spearheaded projects to improve the health of the nation and the world.
- Three thematic collections: the 1964–2000 Reports of the Surgeon General, the history of the Regional Medical Programs, and the Visual Culture and Health Posters.

The site gives researchers, educators, and future scientists all over the world access to unique biomedical information previously accessible only by making in-person visits to the institutions holding the physical manuscript collections. 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, including to individuals with disabilities. The growing *Profiles in Science* digital library provides ongoing opportunities for experimentation in digitization, optical character and 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, or establishes collaborations with external institutions which hold papers of interest. The Web site averages more than 95,000 unique visitors each month, including people seeking an authoritative source of information about current events, such as the February 2013 death of former U.S. Surgeon General C. Everett Koop. When the July 25, 2013, Google home-page doodle featured the late Rosalind Franklin's birthday, *Profiles in Science* received nearly 100,000 unique visitors in a single day.

At the end of FY2013, the 36 publicly available collections contained 27,058 items composed of 142,214 digitized image pages, including transcripts of 10,075 handwritten pages or pages we couldn’t use optical character recognition technology for. In addition to updating the *Profiles in Science* collections during FY2013, we increased the visibility of the *Profiles in Science* digital items through user-configurable software that allows customers to display or hide large thumbnail images. To enhance the interoperability and availability of the *Profiles in Science* data, we added an Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) application programming interface (API). We also improved the reliability and security of the underlying software by enhancing documentation, upgrading open-source software, and replacing modules to eliminate dependence on aging third-party software. We increased compatibility with evolving standards by developing, testing, and documenting alternative digitization protocols for multipage items. We offered the research community insight into the digital library underlying the *Profiles in Science* Web sites by publishing "Improving Software Sustainability: Lessons Learned from Profiles in Science" in the proceeding of the Society for Imaging Science & Technology Archiving 2013 Conference, April 2−5, 2013, and "Why Can't You Just Build It and Leave It Alone?" a guest post for the Library of Congress digital preservation blog, The Signal, on June 5, 2013.

Evidence-Based Medicine: PubMed for Handhelds (PubMed4Hh)

Developed and released in FY2003, PubMed for Handhelds (PubMed4Hh) facilitates evidence-based medical practice with MEDLINE access from almost anywhere via smartphones, wireless tablet devices, netbooks, and portable laptops. PubMed4Hh requires no proprietary software and reformats the screen display as appropriate for the wireless handheld device being used. Clinical filters feature easy access to relevant clinical literature.

Newly developed resources allow people to search MEDLINE through text-messaging. An algorithm to derive “the bottom line” (TBL) of published abstracts allows clinicians to quickly read summaries from almost anywhere. For example, it enables doctors to quickly consult the findings from research to help determine the best course of treatment for a patient. A “consensus abstracts” element provides rapid review of multiple publications with smartphones. A recent review of PubMed4Hh server logs showed that more than 90 percent of queries were clinical in nature.

To evaluate the usefulness of abstracts in clinical decision-making, randomized controlled trials using simulated clinical scenarios were conducted by the Uniformed Services University of the Health Sciences, the Botswana−University of Pennsylvania partnership, and the National Telehealth Center and Philippine General Hospital, Manila. Using simulated clinical cases, these studies demonstrated the usefulness of the app for clinical decision-making.

The PubMed4Hh app is available for iOS (iPhone/iPad) and Android users. Since its release in September 2012, the iOS app has been downloaded more than 138,000 times, by users in the U.S. (44 percent of downloads) and elsewhere (56 percent). Queries from smartphone apps now account for 60 percent of all queries. The total number of searches has tripled since before the introduction of the smartphone app.

Clinical Vocabulary Standards and Associated Tools

Many of our projects in this area continue to promote the development, enhancement**,** and adoption of clinical vocabulary standards. During FY2013, we provided

comments and assistance to ONC, Centers for Medicare and Medicaid Services (CMS), and FDA, often on extremely short deadlines, related to health IT clinical vocabulary and messaging standards for quality measures and meaningful-use regulations. We reviewed several hundred quality measures proposed for phase three meaningful-use regulations and gave high-level feedback to CMS related to their terminology, clinical validity, and compatibility with clinical workflows. We also reviewed and provided comments about:

- Electronic messaging in response to draft implementation guides issued by the standards development organization Health Level Seven, Inc. (HL7).
- Templates for reporting the results of laboratory testing for cancer biomarkers generated by the College of American Pathology (CAP).

The CORE Problem List Subset of SNOMED CT

SNOMED CT is a comprehensive, multilingual medical terminology for anatomic sites, organisms, chemicals, diagnoses, symptoms, findings, and other such concepts. The problem list — a patient's list of active conditions and symptoms — is an essential part of the electronic health record (EHR). Meaningful-use regulations from CMS require the use of SNOMED CT to code the problem list and many other EHR fields.

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

Since its first publication in 2009, the CORE Subset has received considerable attention from the IHTSDO (International Health Terminology Standards Development Organization), the SNOMED CT user community, EHR software vendors, and terminology researchers. It has been installed in various EHR products and used as a focus for SNOMED CT-related research, mapping projects, and quality assurance. The MedlinePlus Connect Project, which facilitates linkage by medical records systems and other outside sources to NLM's rich consumer sources of medical information, has mapped the CORE Subset to MedlinePlus health topics so that medical records systems could automatically pull this educational material for patients to use. In 2012, the CORE Subset was enriched with a clinical dataset from the US Department of Veterans Affairs covering three million patients.

The CORE Problem List Subset of SNOMED CT currently contains about 6,000 concepts and is published in the UMLS as a specific content view. The CORE Subset is updated four times a year to synchronize with changes in SNOMED CT and the UMLS.

Mapping between SNOMED CT and ICD codes

International Classification of Diseases (ICD) codes are required for public health reporting of population morbidity and mortality statistics. In the US, ICD-9-*CM* (the ninth version "Clinical Modification") is also used for reimbursement, to be replaced by ICD-10_CM in the future. Because of this need, many existing EHR systems are still using ICD-based vocabularies to encode clinical data. However, ICD was not designed to capture information that is detailed enough to support clinical care. SNOMED CT is a much better clinical terminology for that purpose, and its use will be required as part of the meaningful-use regulations.

To encourage the migration to SNOMED CT, and to enable EHRs to output ICD codes for administrative purposes, we have developed various maps between SNOMED CT and the ICD classifications. We published a SNOMED CT to ICD-10-*CM* rule-based map, covering 35,000 SNOMED CT concepts. This map allows users to encode patient problems in SNOMED CT terms, and then to generate the appropriate ICD-10-*CM* codes in real time for billing or other purposes. To demonstrate the use of the map, we developed the I-MAGIC (Interactive Map-Assisted Generation of ICD Codes) demo tool.

For an international project, in collaboration with the IHTSDO and the World Health Organization (WHO), we assisted in the development of an analogous rule-based map from SNOMED CT to ICD-10. We adapted our I-MAGIC tool to showcase this map to ICD-10 as well. In a separate project, to help convert legacy ICD-9-*CM* encoded clinical data into SNOMED CT codes, we produced two more maps to SNOMED CT from 9,000 commonly used ICD-9-*CM* diagnostic codes, and 3,000 ICD-9-*CM* procedure codes.

RxTerms

RxTerms is a free, user-friendly, efficient drug-interface terminology that links directly to RxNorm, the national terminology standard for clinical drugs. The Centers for Medicare and Medicaid Services (CMS) used RxTerms in one of their pilot projects in the post-acute care environment. RxTerms is also used in the NLM PHR and at least one EHR from a major medical institution in Boston. During FY2012, we aligned the data model of RxTerms and RxNorm by creating a new term type in RxNorm to cover the drug-route combination. We're continuing to align data elements between RxTerms and RxNorm, and we're reviewing the dose-form information in RxTerms to improve usability. We update RxTerms monthly and bundle it with the full release of RxNorm.

RxNav

When we released RxNav in September 2004, it was a graphical interface browser to the RxNorm database and was primarily designed for displaying relations among drug entities. We then integrated two other drug information sources into RxNav: RxTerms, an interface terminology for prescription writing or medication history recording, and NDF-RT (National Drug File - Reference Terminology (NDF-RT) is produced by the U.S. Department of Veterans Affairs), a resource that links drugs to their pharmacologic classes and properties, including indications, contra-indications, and drug-drug interactions. Later, we created SOAP-based and RESTful application programming interfaces (APIs) so system developers and researchers can easily integrate RxNorm functions into their applications. Examples of such functions include mapping drug names to RxNorm, finding the ingredient(s) corresponding to a brand name, and obtaining the list of National Drug Codes (NDCs) for a given drug.

During FY2013, we expanded RxNav to include additional drug information sources: drug classes from the Medical Subject Headings (MeSH) and the Anatomical Therapeutic Chemical (ATC) classification of drugs (developed by the World Health Organization (WHO) and widely used in Europe).We also released RxMix in FY2013, a graphical interface allowing users to create complex queries to drug information sources (that is, complex sequences of API functions) and to execute them on single values or on a list of values, in batches. For example, users could generate a list of statin drugs, or drugs containing a particular ingredient (such as penicillin), by identifying drugs in RxNorm that have a given property in NDF-RT. This ability enables applications and electronic medical records to create lists that can be used to help automate critical tasks, such as avoiding prescribing medications to which a patient has known allergies and other clinical decision support.

Usage of RxNav, and the SOAP (Simple Object Access Protocol XML-based information exchange protocol) and RESTful (Representational State Transfer system architecture) APIs for RxNorm, RxTerms, and NDF-RT, received a combined total of about 100 million queries during FY2013, double the number in 2012. In FY2013, we presented these APIs at the HHS Health Data Initiative. Users include clinical and academic institutions, pharmacy management companies, health insurance companies, EHR vendors, and drug information providers. Developers of mobile apps have also started to integrate our APIs into their applications.

LOINC Standards for Identifying Clinical Observations and Orders

Within medical record systems, patient summaries, and reports to public health organizations, Federal Meaningful Use 2 EHR regulations require lab result messages sent to ordering clinicians to use LOINC (Logical Observation Identifiers, Names, and Codes). In FY2013, we continued to work with the Regenstrief Institute, major laboratory companies, several NIH institutes, and other organizations to expand the size and breadth of the LOINC database.

By the end of FY2013, LOINC had more than 26,000 users in 157 countries and had been translated into 12 languages, including a new Russian translation. The Regenstrief Institute and the International Health Terminology Standards Development Organization (IHTSDO) signed a long-term agreement to begin collaborative work on linking their leading global healthcare terminologies: LOINC and SNOMED Clinical Terms.

We worked with Regenstrief and the LOINC Committee to create more than 1,700 new LOINC terms for both laboratory and clinical variables, and the LOINC database now contains more than 72,000 terms. We released new terms for radiology, toxicology, chemistry, hematology, molecular pathology, antibiotic susceptibility, the 2003 version of U.S. Standard Birth Certificate and Fetal Death Report panels, newborn hearing and critical congenital heart disease (CCHD) screening, the Neonatal Skin Risk Assessment Scale (NSRAS), and other new survey instruments. During FY2012, we also edited existing molecular genetics terms to harmonize with Human Genome Organization (HUGO), Human Genome Variation Society (HGVS), and International System for Human Cytogenetic Nomenclature (ISCN) recommended nomenclature.

To facilitate electronic reporting of lab results, we worked with four of the eight largest international laboratory instrument vendors to help map or check the mapping of their internal instrument codes to LOINC codes. All eight such vendors now assert they provide LOINC codes for all the test codes their instruments can generate. We also worked with many smaller vendors to find (or create new) LOINC codes to describe the results of their test kits or instruments.. We provided technical advice to the CMS Office of Office of Clinical Standards and Quality on creating standardized variables and data elements for all CMS data-collection tools for post-acute-care assessment.

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

- National Eye Institute (NEI) to restructure its packages of assessment instruments for the National Ophthalmic Disease Genotyping Network (eyeGENE®),
- National Center for Advancing Translational Sciences (NIH/NCATS) Office of Rare Disease to revise the common data elements (CDEs) for their registry system for rare diseases—and we plan to create corresponding LOINC codes,
- National Heart, Lung, and Blood Institute (NHLBI), NICHD, and NIH/NCATS on the NHLBI Hemoglobinopathies Uniform Medical Language Ontology (HUMLO) project, to develop CDEs for hemoglobinopathies using standard terminologies such as LOINC for the questions and SNOMED CT for the answers, and
- National Institute of Neurological Diseases and Stroke (NINDS) on CDEs and the neurological quality-of-life (Neuro-QOL) measures.

Newborn Screening Coding and Terminology Guide

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

During FY2013, we worked closely with many states and vendors on their implementation of newborn screening LOINC and SNOMED CT coding and HL7 messaging standards. We revised many LOINC terms and created several new terms based on requests from NBS programs and laboratories. We created a comprehensive LOINC panel to report screening results for critical congenital heart disease (CCHD), the latest condition added to the HHS Secretary's Recommended Uniform Screening Panel for NBS. We expanded beyond NBS and worked with many states to create a panel of terms for reporting the results of therapeutic diet monitoring for patients with conditions that were diagnosed based on NBS. We also participated as a technical advisor about terminology and interoperability standards for HRSA's CCHD-pilot-program grantees and represent NLM on the Newborn Screening Technical assistance and Evaluation Program (NewSTEPs) Steering Committee and Health Information Technology workgroup.

Communication Infrastructure Research and Tools

We perform and support research to develop and advance infrastructure capabilities such as high-speed networks, nomadic computing, network management, wireless access, security, and privacy.

Videoconferencing and Collaboration

We continue 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 continued monitor the high-definition (HD) open-source work of video conferencing tool (VIC) developers on H.264 compression.

Our team also investigated some new cloud collaboration tools, including proprietary ones that are standards compliant and that emulate a pioneering collaboration model developed by Argonne National Laboratory. Some commercially available cloud technologies lack features required at NLM, are either too complicated or unstable, or have licensing terms that are not cost-effective given the needs of the Library. Currently, there seems to be no compelling rationale to migrate from the H.323 standard videoconferencing appliances we use for NLM program support. The team continued to collaborate with the Rochester Institute of Technology (RIT) and the University of Puerto Rico Medical Campus, however, to test open-source software and commercial cloud technologies for compressed HD videoconferencing based on the H.264 video standard.

We decided to work exclusively with UltraGrid technology for video conferencing because:

- The technology has been enhanced since we started our collaboration with UltraGrid's developers at Masaryk University in Brno, Czech Republic, last year.
- Other uncompressed technologies are receiving less institutional support.
- The research team at Masaryk is very active and shares our research goals and interests.

The team installed an uncompressed video system at the Medical University of South Carolina (MUSC) for a clinical trial. Given the laboratory focus on UltraGrid, iHDTV technology was moved to the clinical setting. The trial's purpose is to study uncompressed video's use as a diagnostic tool. Investigators selected teledermatology as a research focus because previous research has shown it to be particularly difficult to use standard definition video to perform remote dermatological exams. The project is half-way through the data-collection phase.

We continued to work with Specialized Information Services (SIS) on a distance-education outreach program for minority high school students and with the NIH Library to offer National Center for Biotechnology Information (NCBI) database and other bioinformatics training at a distance. In FY2013, we

conducted bioinformatics programs with the Charles R. Drew University of Medicine and Science, the University of Maryland, the University of North Carolina at Chapel Hill, the University of Tennessee at Memphis, the University of Puerto Rico Medical Campus, and Virginia Commonwealth University. One program, involving three of the participating institutions, used cloud-computing resources that reduced genomic-analysis overnight computational times to a matter of minutes. The simultaneous use of the technology by so many students in computer labs at different sites stressed the cloud architecture, but proved feasible. Because travel funds were cut, the team also supported several successful virtual site visits with Regional Medical Libraries to review programs funded under NLM's Regional Medical Library Program. The team published a comparison of the uncompressed video technologies and a re-analysis of findings from previous telemedicine research reviews that identified conditions affecting telemedicine-intervention success.

Disaster Information Management: Lost Person Finder

NLM's increasing interest in mitigating the effects of wide-area disasters led us to develop several information resources and tools. In our Lost Person Finder (LPF) project, we address the problem of how to reunite families separated by mass casualty events. LPF systems combine image-capture, database, and Web technologies, and address both hospital-based and community-wide disaster scenarios.

Web Site and Services

The heart of our system is People Locator® (PL), the main LPF Web site and its MySQL database. We extensively customized the open-source Sahana disaster-management system to create a unified site to hold data from multiple disasters. Missing or found people may be reported to PL by hospital counselors, relief workers, or the public. PL users can report or search for the missing via computer or through mobile apps using Web services. PL is now running on enterprise-level, load-balanced dual systems with failover (to ensure continued availability and operations if one system fails), independent uptime monitoring (to track server availability and performance), and an indexing engine (powered by SOLR software) that can retrieve records quickly.

During 2013, we created a separate demonstration Web site (https://TriageTrak.nlm.nih.gov) so hospitals can report directly or via TriagePic, a reporting app. We designed and implemented new features to improve the chances that search engines will find our site. In FY2013, we also:

- Created a prototype interface for visual search (query by face image) that allows the use of optional metadata (such as gender and age) to reduce the search space, thereby increasing search speed.
- Added a feature to allow reporting missing persons with multiple photos, enabling the PL to incorporate certain face-matching project researcher datasets.
- Added better notification and alerting functions.

In addition to technical improvements, in FY2013 we also secured a formal Authority to Operate (ATO) as required by NIH, and we obtained registered trademarks for People Locator, ReUnite, and TriagePic.

We collaborated with many organizations and agencies interested in disaster mitigation by participating in a FEMA-led webinar and the Missing Persons Community of Interest (MPCI). We also participated in the Google Summer of Code and Google Code-in international student programs, in cooperation with the Sahana Software Foundation, by giving university students all over the world project-related opportunities to learn about and contribute to open-source coding and other tasks.

Deployments

People Locator has been deployed in disasters since the Haiti Earthquake in 2010, as well as in demonstrations and large-scale multi-institutional drills with local Bethesda hospitals. In 2013, PL was deployed for the Boston Marathon Bombing (March), Sichuan Earthquake (April), Uttarakhand India Flooding (June), Acapulco Flooding (September), and Super Typhoon Haiyan (November). In the case of Typhoon Haiyan in the Philippines, more than 100,000 missing-person records were posted to PL (many through Google's Person Finder which is interoperable with PL). Of these records, about 20,000 had photos of missing people.

Mobile Apps

For hospital-based reporting, the triage process begins with TriagePic, a Windows application that hospital staff can use to quickly photograph arriving victims. These pictures, along with general health and triage status and minimal descriptive metadata (e.g., name, age range, gender) are packaged and sent by Web services to PL.

In 2013 we moved TriagePic from a laptop-hosted system to both iOS and Android platforms, adding AES 128 encryption of patient text data and photos to these apps. In addition, we investigated approaches to design a version of ReUnite that supports multiple screen sizes—from cell phone to tablet, from 4-inch screen to 10-inch—to better support the many new mobile devices as the market rapidly expands.

Face-matching Research

Our goal here is to enable users to find missing-person records through automatic face recognition, a significant extension of our current method of searching by name or other text metadata. Our work faces special challenges: unlike many other systems, our face-matching needs to rely on a *single* photo of a person to identify her or his face

in other images, so we can't exploit traditional face-recognition models that require large sets of photos to train the system.

In 2013, we made considerable progress in face localization, a key first step in face matching. Traditional face-localization methods fail by either missing a face or finding too many false positives— that is, objects that the algorithm misidentifies as a face. We minimized these errors and made the algorithm more robust by introducing the simple idea that a face will have skin color. A custom skin-color space was developed using Neural Network–based learners. All pixels in the image that are considered "skin-like" are flagged as potential skin zones, and these zones are contrast enhanced. On reprocessing the image we find fewer false-positive and false-negative errors. In comparison tests with state-of-the-art commercial software, our algorithm performs as well as or better than other methods, not only on high-quality images, but also on the lower-quality images typically sent from disaster sites. In addition, with a view to speeding up face localization, which demands a lot of processing power, we have adapted our algorithm for computation using graphical processing units (GPUs) that offer up to 20X speed up over a CPU alone. The method is currently undergoing testing, and is being readied for deployment in PL.

In addition, we improved face-matching methods by studying several state-of-the-art similarity metrics and implementing two that used scale- and rotation-invariant measures. Further, we developed a novel rotation- and scale-invariant line-based color descriptor. These methods not only improve the system's face-matching accuracy, but also make it robust to variation in illumination. The system was deployed on our staging server and is undergoing testing, and we are working to improve search times.

In an effort to make the face matching robust to variations due to illumination, age, and injury, we're developing face-*region*-based matching algorithms. One that we've developed extracts and matches localized eyes and mouth regions. It performs at 94 percent accuracy but is sensitive to the size of the image. Further improvements to this method are under way.

To support research and testing, we need annotated images. For this purpose, we developed machine-learning tools such as ImageStats to gather objective and accurate "ground truth" data. In 2013, in collaboration with the Sahana Software Foundation, student volunteers from around the world, recruited through the Google Code-in contest, improved ImageStats by contributing crowd-sourced software.

Video Production, Retrieval, and Reuse Project

This development area encompasses projects that contribute to the NLM Long Range Plan goal of promoting health literacy and increasing biomedical understanding. The NLM Media Assets Project gives the NLM easy access to audio-video resources for improved biomedical communications.

Movement Disorders Video Database/MDmedia

This ongoing LHNCBC Research Support Project contributes to improving access to high-quality biomedical-imaging information. We're now able to include video in the clinical study of patients with movement disorders and research the role of mobile technologies in the management of Parkinson's disease by patients and their caregivers.

In FY2013, in collaboration with the National Institute of Neurological Disorders and Stroke and the NIH Movement Disorders Clinic, we developed prototype mobile applications (for iPad and Samsung Galaxy), based on the Movement Disorders Video Database (MDVD), to help diagnose and predict the course of movement disorders. On the basis of input from both clinical and patient-advocacy groups, we also developed two prototype apps for research purposes to measure two functional abilities whose impairment can give insight into the severity of movement disorders: finger-tapping speed and accuracy, and rapid "saccades" eye movement.

A new initiative emerged from our collaboration with clinical and patient groups, to explore the value of video and voice data in mobile devices for patients suffering from movement disorders. Phase one of this research effort focused on interviews with patients at the NIH Clinical Center, soliciting their feedback on how mobile technologies could help them better manage their care between clinician visits. The results of these interviews will be a list of mobile-platform capabilities and associated expected outcomes that patients and clinicians identify as enabling them to manage their care better and improve their overall health. So far, the interviews have suggested that:

- Video and voice records can be valuable.
- Patients could easily use cell phones to share video and voice data with their doctors and have the data included as part of their medical record.
- There is great potential for the use of video and voice journals to monitor and graph symptoms over time.
- Touch-screen technology is more accessible than PC or mouse technology for patients with movement disorders.

Based on these interviews, we will identify viable technologies, then develop and validate these technologies in future studies.

MedlinePlus Interactive Tutorials

In keeping with the LHNCBC mission to develop new technologies and core resources for disseminating biomedical information, we've been engaged in the research and development of augmented-reality modeling and applications, mobile apps, ultra-high-definition

imaging research, image-rich Web sites, and audio/video/imaging archives.

In FY2013, with the NLM MedlinePlus team, we developed a design and navigation prototype for new MedlinePlus Interactive Tutorials to enhance their usefulness. To enable broader public distribution, and allow for an equal experience on all operating systems and devices, we created tutorials with HTML5-compliant animation and interactivity.

Native Voice Mobile Exhibition Planning, Development, and Deployment

In FY2013, NLM published two major LHNCBC-produced Native Voices iPad app updates. They feature improved, enhanced interactive designs and 34 new video clips from interviews conducted by the NLM Director with 15 additional tribal representatives, plus high-resolution images of Native objects and artifacts illustrating the connectedness of wellness and Native culture.

In collaboration with the NLM Office of the Director, the Office of Health Information Programs Development, the Office of Communications and Public Liaison, Specialized Information Services, and the History of Medicine Division, we developed pilot-testing survey measures to collect feedback from the first installation of the traveling version of the exhibition to occur in early FY2014. The pilot-test results will be incorporated into the future planning, development, and deployment of the traveling exhibition.

Computing Resources Projects

We conduct numerous projects to build, administer, support, and maintain an integrated and secure infrastructure to facilitate LHNCBC's research and development (R&D) activities. The integrated secure infrastructure contains network, security, and facility management, as well as system administration support for a large number of individual workstations and shared servers.

In FY2013, as part of Federal Green Information Technology (IT) initiatives, we oversaw continuous monitoring of information systems, and we updated hardware and software to comply with Federal Information Security Management Act (FISMA) 2.0 requirements and NIH guidelines. We achieved a 90 percent FISMA compliance rate (10 percent more than NIH requirements and FY2012 baseline). We reduced purchases of personal IT equipment, purchased and deployed network printers as shared staff resources, and cut locally used device purchases by 80 percent. To avoid interruptions of service, as part of our disaster planning and Continuity of Operations (COOP) risk management efforts, we created a new contact list of Subject Matter Experts (SMEs), continuous monitoring processes, detailed recovery procedures, and centralized data storage. We also upgraded the DMZ (demilitarized zone) perimeter network—a system configuration that provides an additional security buffer—to 10-gigabit capacity with revised Internet Protocol version 6 (IPv6) features (which increase efficiency and network security), and successfully implemented these upgrades without downtime impacts on public service.

The network management includes the planning, implementation, testing, deployment, and operation of high-speed networks over Internet and Internet-2. One core project implements the 10-gigabit network and studies many advanced communication protocols to support LHNCBC collaboration activities and research projects including multicast video conferencing and IPv6. Another core project implements a network monitoring system that displays network usages in real time. The network management team also participated in the NIH-wide 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 a safe IT working environment. One core project studies and implements a unified patch management to improve LHNCBC's overall security measures. Another core project implements the automated security audit system that ensures all systems at LHNCBC comply with policies. The security management team also studies and evaluates the network performance impact of Web anti-virus software, and coordinates annual penetration testing to ensure network security. The NIH penetration test in FY2013 confirmed the effectiveness of LHNCBC's security practices.

The facility management team deploys new IT equipment and servers, including power acquisition, network planning, cabling connection, and space allocation in the central computer room as well as at co-location facilities. Another core project studies, designs, and implements an enterprise console-management system that enables LHNCBC to remotely manage large numbers of servers.

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

Training and Education at LHNCBC

LHNCBC is a major contributor to the training of future scientists and provides training for postdoctoral fellows and other people beginning their biomedical research careers. Our Medical Informatics Training Program (MITP), ranging from a few months to two years or more, is available for visiting scientists, postdoctoral fellows, and graduate and medical students. Each participant spends between a few months and several years working on LHNCBC research projects under the guidance of 16 LHNCBC research staff mentors. The FAQs include information about the lecture series that is part of this program. Participants also make presentations, write manuscripts, attend and present at professional conferences, and may publish in professional journals.

During FY2013, 39 trainees (including 13 postdoctoral fellows) from 15 states and seven foreign countries received training and conducted research across the spectrum of topic areas and disciplines addressed by LHNCBC.

We emphasize diversity by participating in programs for minority students, including the Hispanic Association of Colleges and Universities and the National Association for Equal Opportunity in Higher Education summer internship programs. In FY2013, two students from the Washington Internship for Native Students—both from the Apache nation in Arizona—worked on projects with LHNCBC mentors.

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

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 25 years, the ability to integrate vast amounts of complex and diverse biological information created the scientific discipline of bioinformatics. The flood of genomic data, most notably gene sequence and mapping information, has played a major role in the increased use of bioinformatics. Recently, next-generation sequencing technologies have been a source of large volumes of sequence data. NCBI meets the challenge of collection, organization, storage, analysis, and dissemination of scientific data by designing, developing, and providing the public with the tools, databases, and technologies that will enable genetic discoveries of the 21st century.

NCBI supports a multidisciplinary staff of scientists, postdoctoral fellows, and support personnel. NCBI scientists have backgrounds in medicine, molecular biology, biochemistry, genetics, biophysics, structural biology, computer and information science, and mathematics. These multidisciplinary researchers conduct studies in computational biology and apply the results of their research to the development of public information resources.

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

In order to fulfill its mission, NCBI:

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

Molecular Biology Information Resources

NCBI's molecular biology information resources are based on sequence repositories upon which curated and annotated sets of data resources are built. Information ranges from genetic sequence data to entire genomes, protein sequences and structures to chemical structures and assays, as well as clinical data paired with genotypes. An integral part of the infrastructure underpinning NCBI's molecular biology information is also 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. Together NCBI, EMBL, and DDBJ make up the International Nucleotide Sequence Database Collaboration (INSDC). Other cooperative arrangements, such as those with the U.S. 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 862,350 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 or the Sequence 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 291 million sequence records and over 645 billion basepairs. The traditional nucleotide sequences division increased to 167 million records in FY2013, from 156 million records in FY2012. The WGS division grew to over 124 million records, up from 84 million in FY2012. Six updated releases of GenBank were made public over the past year.

Substantial resources are devoted to the analysis and curation of sequence data. GenBank indexers with specialized training in molecular biology create the records, applying rigorous quality controls. NCBI taxonomists consult on organism classification and, as a final step, senior NCBI scientists review the records for biological accuracy. GenBank now contains more than 1,650 complete genomes from bacteria and archaea, with 25 percent of these deposited during the past year. The number of eukaryote genomes with significant coverage and assembly continues to increase as well, with over 12,722 WGS assemblies now available.

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

GenBank data submission can be accomplished via two resources, BankIt and Sequin. The BankIt submission tool allows the author to submit sequences within an online form and validates submissions by flagging errors before the sequences are deposited. Sequin is a stand-alone tool for updating and submitting large groups of sequences to the database. Improvements to both tools have been made to enhance the user experience. Submission wizards were developed in the NCBI Submission Portal to facilitate the submission of WGS, TSA and their associated BioProjects and BioSamples. An improved annotation service for prokaryote genomes was also released during FY2013.

Genome Information Resources

NCBI plays a key role in the scientific community by 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, individual sequence variations, and genotypic and phenotypic variations. NCBI's genomic information databases include: Assembly, BioProject, Genome, dbSNP, reference sequences (RefSeq), CCDS, dbGaP, Gene, Probe, UniGene, HomoloGene, dbVar, ClinVar, Genome Reference Consortium, GEO, 1000 Genomes browser, CloneDB, Epigenomics, and BioSample. Genomic tools include BLAST, Genome Workbench, Genome Remapping Service, Genome Decoration Page, and Map Viewer. Many of these resources and tools depend on, or otherwise integrate information from, NCBI's RefSeq project.

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 36,036,343 proteins for 31,646 organisms, representing a 100 percent increase in the number of proteins and a 71 percent increase in the number of organisms in the last year. The Microbial section of RefSeq realized significant growth in the number of organisms (97 percent), followed by Fungi (69 percent) and Vertebrate Mammalian (56 percent). The greatest growth in the number of proteins occurred for the Vertebrate Other section (208 percent), followed by the Vertebrate Mammalian section (151 percent), and the Microbial section (110 percent). It is important to note that the growth in Microbial organisms reflects a policy change that all bacterial and archaeal genomes meet minimum assembly quality thresholds, including highly similar sequences that are submitted during disease outbreak monitoring. To prevent a significant increase in redundant proteins in the Microbial section, the policy change was accompanied by the introduction of a new data model for bacterial and archaeal RefSeq proteins such that proteins that are found to be identical on different genomes are now tracked with a single accession, instead of providing an accession for each genome on which the protein is annotated. This policy change for the Microbial section of RefSeq will significantly reduce redundancy amongst Microbial protein records.

Manual curation of data is provided for different aspects of the data for different sections of the RefSeq project. For example, protein names are curated for Microbial RefSeq proteins with results also presented in the Protein Clusters resource. Curation of the transcript and protein sequence, gene structure, and some functional content is provided for the Vertebrate Mammalian and Vertebrate Other sections along with some plants and

invertebrates. This latter category of curation relies on a combination of sequence analysis, publication review, and collaboration. In the past year, the number of organisms supported by the sequence curation group increased by 119 percent to 94 distinct organisms. Curation of individual genes and alternate splice variants resulted in over 31,328 new or updated RefSeq protein-coding, non-coding, and pseudogene records for the 94 organisms supported by the new curation process flow. The record count reflects the pool of RefSeq records (totaling 231,506 transcripts) that are considered well-supported and are available for updates by RefSeq curation staff. RefSeq curation activities resulted in content modification for over 10,000 Gene records and review of over 9,243 RefSeq records to add functional attribute content annotation on the records. Functional content is presented on the sequence records in a "RefSeq-Attribute" comment and consists of a brief description of the attribute (for example, imprinted gene) followed by support evidence for that attribute (a publication, inferred by sequence conservation). In addition, annotation of sequence records was expanded to include reports of RNA-Seq evidence for RefSeq records when applicable.

During fiscal year 2013, RefSeq curators helped define metadata and input sequence data to initiate the NCBI eukaryotic genome annotation pipeline, defined custom BLAST databases, added or updated organism pages for the Genome resource, and performed QA review of annotation result presentation prior to public release in the Map Viewer resource for approximately 50 organisms. RefSeq curators also serve as a point of contact for research communities and 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 of their organisms of interest.

The Consensus Coding Sequence (CCDS) database, a RefSeq sub-project, identifies a core set of high-quality human and mouse protein coding regions that are consistently annotated by both NCBI's eukaryotic genome annotation pipeline and the European Ensembl genome annotation pipeline. Both human and mouse datasets were updated in the past year, yielding current totals of 29,045 CCDS IDs for human and 23,093 CCDS IDs for mouse. The CCDS project is supported by curatorial review so that annotation updates to proteins that have a CCDS ID are done in a synchronized manner that retains consistency in different international genome browsers. Updates are reviewed and must be agreed upon by members of an international collaboration. Public annotation is provided to explain the evidence for the curation decision to make a change. Over the last fiscal year, over 235 existing CCDS entries were subject to collaborative review and 340 explanatory public notes were added to the database. During this time, the CCDS collaboration also worked on reconciling annotation differences for human protein-coding genes that are not yet represented in the CCDS database. Of note, CCDS review often focused on larger genomic regions resulting in collaborative annotation updates of genes not tracked in the CCDS database, such as non-coding RNA loci and pseudogenes.

NCBI's Gene database provides a unified query environment for information about genes and other loci defined by sequence and/or map location. It integrates information about genes and gene features annotated in RefSeq and collaborating model organism databases, and provides a computationally accessible definition of a gene that is extensively integrated into other NCBI resources. Gene is heavily used and includes information for more than 13 million genes from more than 11,000 taxa. About 100,000 full records are retrieved each weekday from 30,000 sessions. In order to maintain a high standard of service, Gene's data flows and functionality undergo extensive testing and quality assurance, and comprehensive documentation is provided to aid public users.

New developments in Gene over the last year included several valuable features: (1) integrating the display of RNA-seq expression data produced by the eukaryotic genome annotation pipeline to support interpretation of the regions of the genome that are expressed; (2) improving representation of human genes that are important in HIV interactions; (3) reporting of additional orthologs; (4) establishing a history of genome annotation; and (5) creating a new tabular report of query results. Some projects are still ongoing.

Human, mouse, and other medically important taxa represent some of the most extensively used parts of Gene, and Gene staff is heavily involved in ensuring high-quality annotation of genes on genomic sequences for these taxa. In FY2013, improvements in the NCBI eukaryotic genome annotation pipeline helped to produce annotations for 58 taxa, many of which have no annotation available in any other public resource. In particular, three full human assemblies were annotated in two different annotation runs in FY2013 in order to maintain a timely and complete representation of human records in Gene.

RefSeqGene defines genomic sequences to be used as reference standards for well-characterized genes. The representation of human genes as RefSeqGene records continued to grow, from 4,849 at the beginning of the fiscal year, to 5,053 at the end. Activity within the Locus Reference Genomic (LRG) collaboration also increased. Currently about 786 RefSeqGenes are assigned LRG accessions.

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

Users are able to zoom into progressively greater levels of detail, down to sequence data for areas of interest.

The Genome Decoration Page allows users to place their own annotation data on graphical ideogram representations of the genomes of nearly 20 organisms. In FY2013, the Bulk Sequence-Cytogenetic Conversion Service was made public. This tool, which can be accessed as a Web interface from the Genome Decoration page or via an API, allows users to get cytogenetic locations for features or sequence locations, or get sequence locations for cytogenetic locations.

The NCBI Remap tool allows users to remap coordinates of genes, SNPs, and other markers from one genome assembly to another. The tool accepts a variety of genome annotation formats and uses the alignments of one assembly to project annotation features on the other assembly. The Remap service includes an API that enables users to programmatically implement remaps. The Clinical Remap tool provides coordinate remapping between genome assemblies and reference standard RefSeqGene records. In FY2013, the Clinical Remap tool added support for mapping of features annotated on Locus Reference Genomic (LRG) sequences, used in the reporting of genomic variants.

The database of Short Genetic Variations (dbSNP) catalogs common genetic variation. dbSNP contains over 232 million submissions of human genome data that have been processed and reduced to a non-redundant set of 62 million refSNP clusters. Approximately 130 other organisms are represented in the SNP database, with 505 million submissions curated to 226 million refSNP clusters. New attributes were added to SNP over the past year for improved searching and filtering of human variation by characteristics including: allele origin, clinical significance, and global minor allele frequency.

The Probe database is a public registry of nucleic acid reagents designed for a wide variety of applications, along with information on reagent distributors, probe effectiveness and computed sequence similarities. 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 15 million probes.

UniVec, NCBI's non-redundant database of vector sequences, is used in conjunction with the VecScreen tool to screen nucleotide sequences for contamination with foreign sequences introduced during the cloning or sequencing process. The UniVec database was updated to build number 7.1 during the year, and the number of sequences represented in the database increased by 3 percent. The newly added sequences include 43 complete vector sequences and 23 adaptor, primer and multiple-cloning site sequences. The new additions include sequences that NCBI staff had observed to contaminate many sequences submitted to GenBank, and also include several sequences suggested by users. This update to the UniVec database will enable searches run using NCBI's VecScreen tool to detect more of the foreign sequences added by various cloning procedures.

The Genome Reference Consortium (GRC) is an international collaboration that aims to update and improve the human, mouse and zebrafish reference genome assemblies. NCBI provides informatics support for the project, such as tracking tiling path files, tracking overlaps between adjacent clones, and curation. NCBI also generates the final assembly after collaboration and quality assurance. In FY2013, the GRC released four non-coordinate changing updates to the human reference genome assembly, adding 52 fix patches and three novel patches, and one to the mouse reference genome assembly. Concomitant with this work, the GRC focused its attention on tasks associated with a major assembly release to the human reference genome made at the end of calendar year 2013. At NCBI, this work involved several sequence analyses and the development of new tools and dataflows to enable large-scale sequence additions to the assembly, including modeled centromeres. In addition, NCBI collaborated with external groups sequencing two other human genomes (CHM1 and RP11), assisting with their assembly production, providing assembly QA and assessing the suitability of selected sequences for inclusion in the updated reference assembly. NCBI also made improvements to the software that maps issues on which the GRC is working to the various assemblies and implemented several GRC Web site updates to improve public access to GRC curation efforts. This year the GRC also began collaborating with the Maize Genome Database to pursue their use of GRC curation tools for updates to the maize genome reference assembly.

Comparative Genome Data

The Genome database provides sequence and map information for a wide variety of organisms including eukaryotes, archaea, viruses, and bacteria. In FY2013, the Genome site was reorganized to represent a rapidly growing number of bacterial genome assemblies. Some bacterial Genome records represent thousands of sequenced and assembled isolates. These genomes are organized in closely related groups calculated by genomic BLAST alignments. Genome overview pages show genome groups in tabular or distance tree formats.

The BioProject database provides access to large-scale biomolecular projects that may include a variety of data types, such as single-cell organism genomes, metagenomes, transcriptome sequencing, genotyping, variation and epigenomics. The resource provides tracking of several data elements, including detailed information about a project’s scope, material, and objectives. In FY2013 over 9,000 bioprojects were created via the Web-based submission system while about 1,000 others were created by GEO and the non-Web-based system.

The Assembly database provides statistics, update history and links to sequences for prokaryotic and eukaryotic genome assemblies. The database tracks

changes that are updated by providing an accession and version number to each assembly. It also contains the placement of each scaffold in the assembly along with the name and sequence accession and version for each chromosome and scaffold. In addition, Assembly organizes and provides assembly metadata such as assembly names, synonyms and statistical reports. Assemblies may also be linked through the Genome database main page or from a genome record.

A new Eukaryotic Annotation Status page was developed this year to inform users on the progress of genome annotation, recently completed annotation, and organism annotation rationale. The new page also includes links to data and Assembly statistics.

Pan-genome Approach

Recent rapid advances in sequencing technologies have provided a relatively cheap and fast way of studying the diversity of microbial species by discovering representatives of novel divisions and analyzing the variation within the species by sequencing closely related genomes from the ecological microbial populations or clinical studies of pathogenic bacteria. Comparative analysis needs a target—a coherent group of individual isolates with some degree of similarity where the degree of similarity is defined by the goal of the study. NCBI developed a robust dissimilarity measure for grouping whole-genome sequence assemblies into clades. The proposed ribosomal protein-marker distance and genomic distance are tailored to achieve robustness while remaining appropriately sensitive. The results are available via the FTP site within genomes and will be regularly updated in coordination with each RefSeq release.

Specialized Databases and Tools

The NCBI Taxonomy Project provides a standard classification system used by the international nucleotide and protein sequence databases. NCBI's Taxonomy database is curated to include the names of species for which sequences have been submitted to the protein and nucleotide databases. Over 30,000 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. A "Common Tree" function builds a tree for a selection of organisms or taxa. In FY2013, Taxonomy staff began curating type material, using the data to label sequences from type specimens or strains in the sequence databases. Type material can range from a preserved specimen in a museum to bacterial strain types as cultures deposited into a culture collection. Sequence from type material is particularly important because when used, species identification is almost certainly correct. A "sequence from type" filter has been added to Entrez queries which the Taxonomy group has used to correct and extend the RefSeq reference set of 16S ribosomal RNA sequences from type strains of prokaryotes to over 10,000 entries. The database also has genomes from type strains for more than 2300 species of prokaryotes, with many more in the pipeline.

The Sequence Read Archive (SRA) was developed to handle large amounts of data generated from massively parallel sequencing experiments. SRA is NIH's primary archive of high-throughput sequencing data and cooperates with the International Nucleotide Sequence Database Collaboration (INSDC) for data exchange. The repository accepts various formats of high-throughput sequencing data, including that produced by platforms from Illumina, Complete Genomics, Applied Biosystems and others. SRA now contains more than a PetaBase of sequence data, including more than 390 TeraBases of open-access data from the three INSDC partners. In addition to raw sequence data, SRA stores alignment information in the form of read placements on reference sequences. An SRA Toolkit containing precomputed binary files for common operating systems and the source code for each release is available for download.

The Clone Database integrates information about genomic clones and libraries, including sequence data, genomic position, and distributor information. CloneDB provides descriptions, sources, and detailed statistics on available genomic libraries for a large number of organisms. In FY2013, the Clone DB search interface was updated to facilitate the identification of clone libraries and individual clones of interest. App-logging queries and reports were also instituted to track Entrez queries and page usage which will be used to guide future updates to the resource. Clone placement maps were newly generated or updated for more than ten organisms, and clone end sequence data were loaded to the database for more than 26 organisms. In addition, the underlying database was updated to better support data associated with distributors of genomic clone libraries.

The Gene Expression Omnibus, or GEO, is a public repository for large-scale functional genomic data generated by microarray- and sequencing-based technologies. GEO DataSets is a study-level database and GEO Profiles is a gene-level database. In FY2013, GEO processed over 10,000 new studies, which represents 20 percent growth over the previous year. An overall increase in curator productivity was largely due to several improvements to the internal curation pipeline, such as enhanced utilities and processes that dynamically balance curator workload. Other GEO resource improvements include implementing faceted querying in both Entrez GEO Profiles and GEO DataSets, indexing additional fields, re-instrumenting GEO pages so that more detailed log analyses may be performed, and modifying the GEO2R analysis tool so that it may be used to examine very large data sets. In addition, the GEO homepage was redesigned and GEO submitter accounts were integrated with MyNCBI accounts.

NCBI's BioSample database contains descriptions of biological source materials used in

experimental assays. Currently, the database contains over 2.3 million records and has experienced 50 percent growth this fiscal year. In FY2013, the BioSample Submission Portal underwent substantial improvements to better capture rich and relevant metadata for an expanded set of sample types. Additionally, the submitter interface was redesigned to better integrate with other Entrez portals, including BioProject and WGS. An XML-based submission pipeline was implemented for automated deposit by large data producers. Major improvements have also been applied to data flow pipelines to ensure that data remain in sync with related resources, including SRA and GenBank.

The NCBI Epigenomics database serves as a comprehensive resource for whole-genome epigenetic datasets selected from the GEO database. Epigenetics is the study of stable and heritable changes in gene expression that occur independently of the primary DNA sequence. The Epigenomics database contains over 4,200 studies, covering more than 1,200 samples from five well-studied species. A new "Upload Tracks" feature was added this year allowing users to view and compare their own datasets with information stored at NCBI. A YouTube tutorial was also released to provide assistance to users.

The NCBI Genome Workbench is an integrated application for visualization and analysis of sequence data. It is designed to provide a flexible platform for development of new analytic and visualization techniques. Four releases were provided to the public in FY2013, supplying users with a number of new features, including: Tools Quick Launch; separation of graphs in sub-tracks; persistent markers on sequences; SNP table view; five-column feature table format reader; text view search; the ability to open projects as tables (allowing users to sort and search project content); publication quality image export (PDF), which allows users to print presentations in high resolution; Load Default Track function, which allows users to restore track configuration; and a new reader for opening text alignment files.

Much of the development and visualization of Genome Workbench is illustrated directly on its Web-based counterpart, Sequence Viewer, NCBI's graphical display for nucleotide and protein sequences. Sequence Viewer is designed to be an embeddable component to compliment other information-rich views. In FY2013, Sequence Viewer version 2.27 was released with new features including tiling path and contig (scaffold) tracks for assembled records, drag and drop reordering, and one-click removal of tracks within the graphical view. Multiple Genome Workbench-related tutorials were added to NCBI's YouTube site this year as well.

The Influenza Virus Resource is a comprehensive collection of flu sequences. Samples collected from around the globe include viruses obtained from birds, pigs, humans, and other species. Links are provided to other flu resources containing sequences, publications, and general flu virus information. More than 61,500 new influenza virus sequences were added to the Influenza Sequence Database in FY2013. About 25,000 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 many other institutions worldwide, and were processed by the NCBI flu annotation pipeline. Flu sequences have been submitted from the current flu season, helping researchers and clinicians understand outbreaks and develop vaccines.

Variation Data Resources

The database of genomic structural variation (dbVar) contains data, including related clinical data, on variant DNA from many organisms. Structural variation data submissions are accepted from whole genome comparative studies and locus- and gene-specific data from quantitative studies. The database, therefore, is organized by study. In FY2013, a dbVar Genome Browser was released, that allows users to browse multiple studies at once and search for any feature annotated on the genome.

The 1000 Genomes Project is an international consortium that aims to build a comprehensive map of human genetic variation. Started in 2008, the project has collected data from 2,600 individuals originating from 25 populations around the world. In addition to cataloging variation data, 1000 Genomes is also a methods development project that is pioneering and evaluating methods for: generating data from next-generation sequencing platforms; exchanging and combining data and analytical methods; discovering and genotyping SNPs and CNVs from next-generation data; and imputing with and from next-generation sequencing data.

The 1000 Genomes Browser shows variants, genotypes, and supporting sequence read alignments from the 1000 Genomes Project. Users may perform searches using gene names, dbSNP accession numbers, or chromosomal positions. The graphic shown within the browser is based on NCBI's Sequence Viewer graphical tool. A table of genotypes organized by the 1000 Genomes Project is shown beneath the graphic view where users can drill down to show individual-level genotypes. The 1000G browser underwent "behind-the-scenes" updates in FY2013 to incorporate and take advantage of updates that were occurring to the NCBI Track Management System (TMS).

The NCBI Track Management System (TMS) is a new, centralized database that manages data for use by various browsers and applications, making it easier for users to find and access data in which they are interested. FY2013 saw implementation of the database and the loading of several hundred new data tracks from a variety of NCBI projects. In addition, TMS integration was added for the Get-RM browser and Variation Reporter.

Variation Reporter provides access to location information by taking a set of locations in a human genome assembly and identifying known human variations at those positions. The variations are collected from the

NCBI refSNP data set within dbSNP. Variation reporter also accepts a variety of genome annotation formats. Query results provide the location of variants along with clinical information, minor allele frequency, links to literature, and functional consequences. In FY2013, the Variation Reporter was updated to include the user uploaded data (UUD) service.

UUD is a new centralized resource that ensures various NCBI tools provide consistent handling of user uploaded data. Data uploaded via UUD is managed by TMS. In FY2013, UUD was also implemented in the Epigenomics browser. Several updates were made in the UUD design in FY2013, which will be publicly implemented in FY2014.

NCBI maintains a Variation Web site that is a portal to NCBI's variation resources and tools, including 1000 Genomes Browser, dbSNP, dbVar, ClinVar, the Variation Reporter, Clinical Remap, and the Phenotype Genotype Integrator. Links for submission queries are also present. FY2013 saw the development of the Variation Viewer browser. This browser, which also uses TMS and UUD, allows users to view, search, and navigate variations housed in dbSNP, dbVar, and ClinVar in genomic context. Users can search based on chromosomal location, gene, variant IDs from dbSNP and dbVar, or phenotype annotated in Gene, and view results in genomic text in the dynamic sequence viewer and annotated tables of variations.

GeT-RM is a Centers for Disease Control and Prevention (CDC) project designed to provide genetic testing reference materials, quality control measures, and proficiency testing for clinical testing laboratories using next-generation sequencing technologies. NCBI has collaborated with the CDC to collect variant call datasets for the HapMap samples NA12878 and NA19240 from several clinical testing and research laboratories, and to provide access to the data. The data is available in the GeT-RM browser which displays the variants in genomic context along with genotype information. Data can also be downloaded from the ftp site. During the fiscal year, 53 datasets were collected from 12 contributing labs, including data from targeted gene panels, whole exome sequencing, and whole genome sequencing.

Other Resources

ClinVar is a recently released resource that archives reports of relationships between human variation and observed health status. Launched with about 30,000 records in April 2013, ClinVar records had almost doubled to over 52,000 by the end of the fiscal year. ClinVar currently has more than 1000 users on an average work day, from about 400 sessions. This resource is under active development. As a participant in the NIH ClinGen project funded just before the end of the fiscal year, ClinVar is a key component of the subset of NCBI's resources focused on medical genetics.

The Genetic Testing Registry (GTR) was developed as a centralized resource for test providers to submit information voluntarily about genetic tests for inherited and somatic genetic variations. GTR includes detailed information about available genetic tests, including their purpose, methodology, validity, and evidence of the test's usefulness, as well as laboratory contacts and credentials. GTR also provides access to information from other relevant resources, such as PubMed, GeneReviews, molecular resources and professional practice guidelines.

Participation in GTR by laboratories has been strong. As of September 30, 2013, GTR had 367 registered laboratories from 39 countries who submitted detailed data on 12,158 tests for 3,640 conditions. Enhancements to data submission utilities were implemented including semi-automated submission of minimally required fields via spreadsheets, supplementation of data previously submitted to NCBI for the GeneTests Laboratory Directory, and user group support within organizations. Along with staff support via webinars, YouTube videos and Web meeting tutorials, the enhancements fostered submission of a large number of tests and multiplex panels; for example there are more than 300 registered tests that evaluate five or more genes or chromosomal locations. Also within the year, the scope of test types widened from clinical tests to tests for cancer and other somatic (non-inherited) genetic variation, as well as research tests offered within the context of a research study.

During FY2013, GTR usage changed from an exploratory usage pattern to one of steady, recurring use similar to that of other regularly used NCBI Web resources. Internal sites (e.g., NLM's Genetics Home Reference), other federal Web sites (e.g. CDC) and external sites (e.g. OMIM) began referring to GTR. In response to a licensing agreement between NLM and the American Medical Association, GTR now features molecular pathology CPT codes.

MedGen, a companion resource to GTR and ClinVar, is NCBI's portal to information about human disorders and other phenotypes with a genetic component. MedGen aggregates terms into specific concepts and assigns identifiers to those concepts, thereby allowing computational access to phenotypic information, an essential requirement for the large-scale analysis of genomic data. MedGen maintains up-to-date information that harmonizes terminology from multiple sources, including the UMLS Metathesaurus, and provides an interface to use those harmonized terms to identify related information in the public arena.

In FY2013, 170,000 concepts were integrated into MedGen. Updated data elements include: term definitions, acronyms, synonyms, hierarchies, semantic type and sources; clinical features, gene, cytogenetic location, variations, RefSeqGene and mode of inheritance; links to published literature including professional guidelines, PubMed, structured queries and reviews; links to other

NCBI resources; links to consumer resources and ClinicalTrials.gov; and information about genetic tests in GTR. Key relationships are maintained and updated including disorders and clinical features as well as those between genes and disorders. Integrated into the Entrez system, MedGen data can also be accessed via ftp site or accessed programmatically via E-utilities.

Chemical Information

PubChem is organized as three inter-linked databases: Substance, Compound, and BioAssay. Together, they form an extensive resource for information on millions of chemical substances. Launched in 2004, PubChem is becoming a key information resource for a number of research fields such as cheminformatics, chemical biology, medicinal chemistry, drug discovery, and RNAi research.

PubChem reached significant growth milestones in 2013 including: over 119 million contributed substance records, over 47 million unique compound records, 715,000 bioassay records, and over 230 data contributors. The PubChem BioAssay database now contains more than 217 million bioactivity test results from over: 2.8 million tested chemical substances; 180,000 RNAi reagents; 6,000 protein targets; and 30,000 gene targets. PubChem now routinely receives more than one million requests per day from more than 120,000 users.

Data submissions from new contributors broadened the scope of PubChem's data contents and expanded public access to these data. Most notably, SureChem (MacMillan) deposited its complete collection of more than 8 million chemical structures extracted from U.S., European, and World Intellectual Property Organization patent documents published since 1976. Over four million of these structures were new to the PubChem Compound database. In addition, thousands of carbohydrate structures were added to PubChem, thanks to contributions from UniCarbKB, an initiative to create open-access information storage and a search platform for glycomics and glycobiology research.

To help support the strong growth of PubChem, a new data submission system, called PubChem Upload, was released to ease and simplify submissions, especially more complicated bioassay data submissions. PubChem Upload offers streamlined procedures for data submissions and includes substantial user interface enhancements, such as an extensive set of wizards, inline help tips, and templates. PubChem Upload replaces the now deprecated PubChem Deposition Gateway, which will be retired in 2014.

In the past year, PubChem released several new data interfaces, such as PubChem Widgets, a Classification Browser, and BioActivity DataDicer. PubChem Widgets provide data interfaces and JavaScript-based display components for some commonly requested data views, such as tabular summaries of patents, bioactivities and PubMed articles linked to a PubChem record. Widgets can also be used to display a carousel of related chemical structures in an animated way, to show the classifications available for a PubChem record of interest, and to provide an autocomplete function that suggests a list of terms when typing in a search field. The new PubChem Classification Browser allows the user to navigate or search PubChem records based on a hierarchical classification system of interest (such as MeSH), providing a powerful way to quickly and visually find a desired subset of PubChem records.

The PubChem BioActivity DataDicer interface allows users to search the entire set of 217 million bioactivity outcomes from the PubChem BioAssay database. Users can query by molecular descriptors, genes, pathways, bioassay categories and other data types, in order to retrieve bioactivity outcomes for genes/proteins that were the targets of bioassays, or for small molecules that were screened in the bioassays. This new interface allows users to traverse, subset, and download the associated genes, proteins, chemical structures, bioactivity data, and more.

As a means to reach out and engage with the scientific community, the PubChem Social Media campaign was launched in FY2013. PubChem accounts on Twitter, Facebook and Google+ were established to share information about new PubChem features and data updates. In addition, the PubChem Blog was started to help inform users of different capabilities of the PubChem system. These Social Media postings augment the existing PubChem announcements page and RSS feed.

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) contains experimentally determined biopolymer structures that have been submitted to the Protein Data Bank (PDB). MMDB is augmented with annotation of structural domains, uniform secondary structures, and explicit links to relevant literature. MMDB is a source of protein and nucleotide sequences, chemicals tracked by PubChem, and novel conserved domains in the CDD, as well as structural neighbors computed by the VAST and VAST+ algorithms on compact structural domains and macromolecular complex structures, respectively. MMDB is processed and released on a weekly basis, following weekly updates of the PDB archive. MMDB is used to provide annotations and homologous 3D structure data via the CBLAST service for the NCBI Entrez Protein database and search results from the BLAST service through detected sequence homology.

In FY2013, MMDB staff finalized work on processing and presenting "merged" structure records.

Limitations of the (soon to be obsolete) data exchange format had forced PDB to break up records describing large macromolecular complexes into several "split" data files. Those "split" records are now merged back together during processing of the structure data and presented as intended by the original data depositors. The merged files make it possible to interactively view and/or download large macromolecular structures in their entirety, such as viral capsids, the rat liver vault, and ribosome structures, all of which could previously only be viewed as structure fragments.

MMDB staff developed VAST+, an extension to the VAST structure neighboring service that reports similarities between macromolecular structural complexes, as opposed to single molecules or pieces of single molecules. VAST+ is the first neighboring service that provides a comprehensive database of structurally similar biological assemblies tracked in the Protein Data Bank and MMDB.

Conserved Domain Database (CDD)

The Conserved Domain Database (CDD) provides information for annotation of functional units in proteins. Two CDD releases were processed and published in FY2013 (versions 3.09 and 3.10). These included a mirror of release 13 of the TIGRFAM database, and 1,146 new or updated NCBI-curated models. CDD curators added 3,355 curated or validated models into the in-house tracking database.

Changes to the RPS-BLAST search algorithm were incorporated into the CDD data production pipeline during this fiscal year. RPS-BLAST now supports composition-corrected scoring, which no longer necessitates masking of compositionally biased query sequences. This in turn increases the overall sensitivity of sequence annotation obtained via RPS-BLAST. Changes in RPS-BLAST behavior triggered changes in the processing of the CDART database, which stores pre-computed sequence annotation for NCBI's Entrez Protein database.

The CDD group also worked with Dr. Andrew Neuwald from the University of Maryland on automating protein domain classification via an algorithm that detects subfamilies conserved in molecular evolution. The algorithm works by sampling subsets from a multiple sequence alignment that conserves sequence signatures that distinguish the domain from the background. This procedure has been found to generate hierarchical classifications that agree, in general, with classifications obtained via manual curation. Work in FY2013 also resulted in adapting the subfamily sampler to the needs of the CDD curation project, and demonstrated the need for further sequence alignment refinement, which can be automated as well. Research undertaken by a summer student demonstrated the potential for significantly speeding up RPS-BLAST searches against the Conserved Domain Database by adapting the BLAST word score threshold to properties of individual domain families. Also, the CDD curation group started work on a set of models that represent repetitive structural motifs and will increase coverage and accuracy of sequence annotation.

In FY2013, the CDD group continued to increase coverage of protein 3D structures with conserved domain models by characterizing new protein domain superfamilies that are not yet modeled by CDD's source databases such as Pfam or TIGRFAMs. CDD was presented at various professional association meetings this year.

NCBI BioSystems Resource

NCBI's BioSystems database was developed as a complement to other databases, such as those that show the components and products of biological pathways along with corresponding annotations and links to the literature. BioSystems contains several types of records, including pathways, structural complexes, and functional sets, and is designed to accommodate other types of information as it becomes available. The BioSystems database serves as a centralized data repository, connects biosystem records with associated literature, molecular, and chemical data through the Entrez System, and facilitates computation on biosystems data.

During FY2013, the BioSystems database incorporated 125,895 new records and had over 1,000 user sessions per day.

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 contains 15 specialized pages corresponding to NCBI databases. During FY2013, BLAST released a new results report that provides faster results, new download options such as the ability to view only aligned regions, a customizable description table, and the ability to display query-based or subject-based views of results in the graphical sequence viewer. Early in FY2013, the nucleotide collection (nt) became the default nucleotide search database for BLAST. Microbial BLAST was improved with new and more effective search options. For nucleotide searches, users can now choose between searching a smaller, curated set of "Representative Genomes" or "All Genomes." Representative Genomes are the best representation of a genome, for a given organism,

chosen by the research community and NCBI computational processes. The All Genomes option provides a choice of complete genomes, draft genomes, and/or complete plasmids. The report offers a new "Genome" link that goes to the Entrez Genomes species page.

Stand-alone BLAST, released in Version 2.2.28 this year, provides a platform for users wanting to perform searches locally. SRA-BLAST was upgraded this year to make it more useful for searching through the more than 700 trillion bases currently housed in NCBI's Sequence Read Archive (SRA).

dbGaP: Integration of Clinical and Genetic Data

The database of Genotypes and Phenotypes, dbGaP, was originally built to house, display and distribute data produced in Genome Wide Association Studies (GWAS). In recent years, the primary mission of dbGaP has expanded beyond GWAS to represent evolving technologies and data types. The data that make up dbGaP studies can typically be categorized into three main classes; 1) Phenotype data, consisting of clinical, anthropomorphic, demographic and exposure variables collected from a few hundred to thousands of individuals; 2) Molecular data produced using SNP chip arrays, expression arrays, epigenetic assays and/or sequencing of RNA or DNA; and 3) Study Documents. Study Documents are protocols or data collection instruments that describe how data or samples were collected and/or processed. The system has proven that it is readily adaptable to house data for many different types of studies that include human phenotype information.

Study Submissions

dbGaP has released 384 studies since 2007, 79 of which were released in FY2013. In addition to new studies, 49 previously released studies were updated with additional data (new versions) in FY2013. The 79 new studies are comprised of phenotype and genotype data from 144,080 additional study participants. The total number of study participants in dbGaP at the end of FY2013 was 577,871.

Linking Studies to Other NCBI Resources

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

Cumulative Summary Counts

Collectively, the studies released include measurements for a total of 577,871 research participants. The database includes:

- Over 146,059 individual-level phenotype and exposure variables.
- 3,373 documents, which allow users to explore the studies through natural language descriptions, collection forms, or scientific protocols, and also provide models for future research.
- Tens of trillions of individual genotypes, i.e., single-nucleotide measurements of participant DNA sequence. These potentially reveal systematic and heritable genetic differences between affected and unaffected individuals.
- 3,810 files of phenotypic association results have been submitted, and each file contains at least 100 thousand p-values. The p-values are test statistics to determine the association between the tested phenotype trait and the variant positions (SNPs) in the human genome.

Authorized Access System Download Activity

The dbGaP authorized access system is the NCBI portal where Principal Investigators (PI) request access for individual-level data housed in dbGaP. By the end of FY2013, 4,133 research projects from 2,525 Primary Investigators had been created in the approval system. Each description of a proposed research activity is prepared by a PI as context and justification for their access to individual-level data. A project may include multiple study/consent group datasets, and a Data Access Request (DAR) is created by the system for each dataset.

Data Usability: Tools and Software Development

During 2013, there was significant progress in the development of the dbGaP submission portal that is currently in use. The improvement will reduce the effort required for investigators and their teams to submit data to dbGaP. The portal will also greatly reduce manual processing of incoming data files. Enhancements to the controlled access and study registration systems were also significant in reducing the time needed to initiate studies and process data access requests.

Entrez Retrieval System

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

molecular sequences via "neighbors" and links provides an efficient and intuitive way of accessing data. Entrez currently supports and integrates 40 databases, including the sequence databases, chemicals, genomes, and biomedical literature.

Literature Information Resources

PubMed

PubMed provides Web-based access to citations and abstracts for the biomedical science journal literature. PubMed is comprised primarily of journals indexed in NLM's MEDLINE database, but also contains a limited number of journals outside the scope of MEDLINE. Links are provided to the full text of articles, when available, via NCBI's PubMed Central database or the originating journal. PubMed, which serves as the foundation of NCBI's bibliographic information system, contains over 23 million citations from more than 36,772 journals, some dating back to the 1800s.

PubMed is continually updated and enhanced for better functionality and more precise search results. During FY2013, the search bar pull-down menu was updated to list the four most recently searched databases. A "Download History" link was added to the History feature in advanced search pages. PM began displaying some publisher-supplied information such as author keywords and non-English abstracts. The NLM Catalog Limits page was replaced with the results filter sidebar.

My NCBI allows users to store searches and results in Entrez with features such as automatic updates and e-mailed results. This year, the "Saved Search" page was modified to include the ability to edit search terms and test the search before saving to My NCBI. SciENcv is a new My NCBI feature that helps users create an online professional profile that may include research activities, honors, and professional contributions, with the option to make it public and share with others. SciENcv complements the My Bibliography service where users can save a record of their own citations (journal articles, books/chapters, patents, presentations and meetings).

PubMed Health

The PubMed Health collection includes systematic reviews of clinical effectiveness published in the last ten years. PubMed Health has close to comprehensive coverage of reliable systematic reviews, combining reviews from leading health technology assessment agencies (such as the US Agency for Healthcare Research and Quality) and the Cochrane Collaboration with the systematic reviews incorporated in the Database of Reviews of Effects (DARE) from the Centre for Reviews and Dissemination (CRD). The Canadian Agency for Drugs and Technologies in Health (CADTH) became a systematic review partner this year.

New systematic reviews are added to PubMed Health at an annual rate of around 7,000, with hundreds of updated systematic reviews added as well. This year, the database added close to 1,800 new or updated summaries or articles for consumers and clinicians based on systematic reviews, including 150 from the National Cancer Institute (NCI), which is providing evidence-based summaries for consumers and clinicians from its Physician Data Query (PDQ) service. By the end of FY2013, PubMed Health contained over 28,000 reviews.

The development of technical infrastructure to build an improved medical encyclopedia and dictionary information for consumers recently began. This includes an anatomical image database, with an interface for anatomy and physiology-based topic pages. Relationships and linkages between, for example, anatomy and related conditions have been enabled. PubMed's filter system was adapted for PubMed Health and implemented this year as well. This process entailed developing a broad range of filtering options, and a guide to filtering searches on PubMed Health. Navigation improvements were also made to the Web site.

PubMed Central

PubMed Central (PMC) archives, indexes, and provides free and unrestricted access to full-text articles from biomedical and life science journals. This repository is integrated with the PubMed biomedical literature database of indexed citations and abstracts. Use of PMC continues to increase in concert with the growth of available articles. As of September 2013, there were over 2.8 million articles in PMC and on a typical weekday 850,000 or more unique users retrieved over 1.6 million articles.

At the end of FY2013 PMC had approximately 1,600 participating journals that deposit either all of their content or the subset of NIH-funded articles they publish. For publication year 2012, PMC has approximately 265,000 articles, about a third of which are associated with NIH funding (extramural research awards or intramural staff authors). About 44 percent of the NIH-funded articles came from the journals that deposit final published articles directly into PMC; the remainder came as author manuscripts via the NIH Manuscript Submission System.

In December 2012, NCBI introduced a new way of displaying articles for online reading. Known as PubReader®, it is particularly well suited for tablets and mobile devices, but also offers advantages to users of traditional desktop and laptop computers. PubReader breaks articles into a traditional multi-column and page format, and includes an image strip at the bottom of the page to easily view figures and tables. PubReader improves article readability and navigation using the latest version of the underlying technologies, HTML5 and CSS3. The software behind PubReader is available online for anyone to use. In September, a new version (1.2) was released that includes a "search this page" feature.

LinkOut

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

In FY2013, the number of organizations participating in LinkOut increased to close to 3,800, representing about a six percent growth rate over the prior year. LinkOut participants include 3,000 libraries, over 520 full-text providers, and 280 providers of non-bibliographic resources, such as biological and chemical databases. Participation in Outside Tool, a service linking to external resources, increased to over 1100 institutions. LinkOut users can now link to 55 million Entrez records, including links to the full text of 61 percent of PubMed records from over 9,700 journals. Usage of LinkOut resources is over 39 million hits per month, and about 1.5 million hits per weekday.

A large scale automatic link check was conducted this year as a quality assurance measure, resulting in a dramatic reduction in the number of broken links and an improved user experience. LinkOut indexing was migrated to CIDX and fine-tuned to work efficiently. The utility for full-text and non-bibliographic providers has been designed and is currently under development and will remain the main focus in the coming year. When it is in place, it will reduce redundant submissions, allow providers to send links using a utility rather than XML files, and provide a clear picture of existing link patterns for better quality control.

E-Utilities

Entrez Programming Utilities (E-Utilities) are a set of server-side programs that utilize 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.

The E-Utilities SOAP schema was updated during FY2013. Users can now search a collection of BLAST databases (db=blastdbinfo) and retrieve metadata and paths to the databases for use in Standalone BLAST. The PubMed E-Utility 2013 DTD update was released in December.

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 NCBI databases and software applications.

CBB research has strengthened NCBI resources 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 100 investigators, staff scientists, research fellows, postdoctoral fellows, and students.

CBB is carrying out basic research on over 25 projects reported in the NIH Intramural Program annual reports of research. The 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 study 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 partly evidenced by the number of peer-reviewed publications generated—over 150 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.

NCBI's Board of Scientific Counselors (BSC), comprised of extramural scientists, meets twice a year to review the research and development activities of NCBI and the research programs of senior investigators in the CBB. The BSC's 41st meeting was held in November 2013.

Bioinformatics Training and Support

Outreach and Education

NCBI's outreach and public services component is essential to ensure that the user community is aware of 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 to anticipate future requirements.

The public services division provides user support via e-mail, by staffing conference exhibits, and through training materials and seminars based on NCBI resources. Over the past year, NCBI staff exhibited at three scientific conferences, presented at seminars and workshops, provided a number of training courses, and published and distributed various forms of tutorial materials. NCBI staff published six articles on NCBI resources in the 2013 Database Issue of *Nucleic Acids Research*. In addition, several resource articles were published in other scientific journals regarding dbVar, PubChem 3D, Clone DB, Genetic Testing Registry, PubMed, and text mining.

NCBI's social media presence continues to increase as usage of NCBI's Facebook, Twitter, and YouTube sites increases. The amount of educational materials published online has also increased and been improved including a newly designed Educational Resources Web page providing links to documentation, courses and workshops, news and updates, and community resources. NCBI added eight tutorial videos to its YouTube site this year including:

- A Guide to NCBI: Gene Expression (3 parts)
- Epigenomics: How to Upload Data Tracks
- Save Searches and Set E-mail Alerts
- My Bibliography: Public Access Compliance
- Need the Full Text Article?
- BLAST: New Results Format
- Sequence Viewer: PDF Rendering Feature
- Using MeSH to Build a Better PubMed Query
-

A new *NCBI Insights* blog was created in FY2013, and includes four categorized types of information: new features, quick tips and tricks for NCBI resources, in-depth explanations of resources, and general scientific features. An NCBI News site was released that will contain announcements addressing changes, updates, and improvements to NCBI Services. The News site will replace the *NCBI News* newsletter that was published as a Web document.

NCBI provides 17 "Announce" e-mail lists that give users the opportunity to receive information on new and updated services and resources from NCBI. Thirty-six RSS Web feeds are available for updates and announcements on specific NCBI resources.

Training

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

NCBI Discovery Workshops take place three times a year on the NIH campus. These two-day workshops, taught by service desk staff, incorporate seminar and hands-on instruction and focus on a variety of NCBI topics: Sequences, Genomes, and Maps; Proteins, Domains, and Structures; NCBI BLAST Services; and Human Variation and Disease Genes. Additional workshops are provided at other locations with assistance from the National Network of Libraries of Medicine (NNLM). In April 2013, a new course entitled, *A Librarian's Guide to NCBI* was provided in conjunction with Library Operations (LO) and NNLM. This off-site course, which will be offered annually, prepares health science librarians for supporting and training their patrons about NCBI databases and tools.

Various workshops and posters were presented at association meetings and annual conferences for biomedical researchers and genetic professionals.

Web Log Analysis

NCBI continues to use Web log analysis and usability studies to improve public Web pages and provide information based on user needs. The NCBI AppLog (application log) system monitors over 4,000 queries on the wide array of NCBI resources, and functionality of the system is continuously improved upon. Recently, improved tracking of user scrolling behavior on Web pages was added which will help reveal the portions of pages of most interest to our users. The accuracy of methods that measure the exact path a user takes through NCBI Web pages was also improved. This has provided an understanding of the links that most successfully achieve their designed purpose. Analyzing user queries allows NCBI to improve and add query sensors that detect particular query patterns, and respond accordingly. Examples include a citation sensor that responds to citation data in PubMed queries and a gene sensor in the sequence databases that responds to gene symbols.

NCBI also continues to use Web analytics to improve the Web user interfaces for major NCBI resources. By using information about how search limits have been used in the sequence databases, a series of sidebar filters was added to the protein database (as was

previously done in PubMed), thereby better exposing these search controls to our users. In addition, based on user query data, the interface for the BLAST page for the Sequence Read Archive (SRA) was improved to include an autocomplete feature that assists users in selecting the desired search database.

EXTRAMURAL PROGRAMS

Valerie Florance, PhD
Associate Director

The Extramural Programs Division (EP) administers extramural grant programs for NLM as authorized by the Medical Library Assistance Act (MLAA) and Public Health Service Act. EP's first grant awards were issued in 1965. 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 opportunity announcements (FOAs) or through special FOAs issued by EP. Each year, NLM makes new and/or continuing awards in these five grant categories: Research Projects, Resources, Career Development, Research Career Training, and Small Business Research & Development.

Overview of FY2013

NLM's EP FY2013 base budget for grant awards was $41,183,998, a 7 percent decrease from FY2012 due to sequestration. Grants award activities were affected by late passage of the FY2013 continuing resolution. Sequestration's impact was felt throughout NIH; Figure 1 shows the effects for NIH extramural programs as a whole, and for NLM's grant programs.

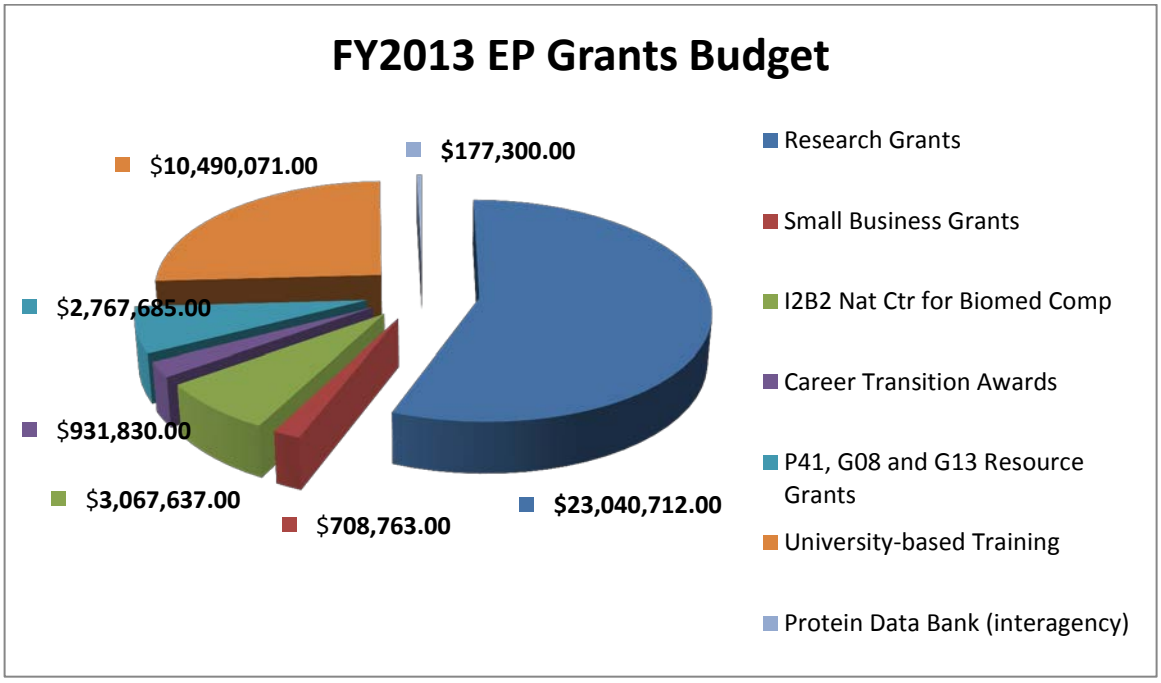
One hundred thirty-seven new and continuing awards were made with NLM's appropriated funds, including 11 co-funds made by NLM to grants administered by other Institutes. Funding in the amount of $2.99 million was received from other NIH organizations for eight NLM grants. FY2013 funds were awarded to 66 universities, colleges, hospitals and small businesses in 31 states, the District of Columbia and the United Kingdom.

Figure 1: FY2013 Sequestration Effects on New Research Project Grants

	NIH	% change from 2012	NLM	% change from 2012
Success rate	16.8%	-0.8%	12.3%	-0.5%
Average size	$ 441,404	-3.0%	$ 383,177	5.0%
Applications received	61,013	-4.0%	103	-33.0%

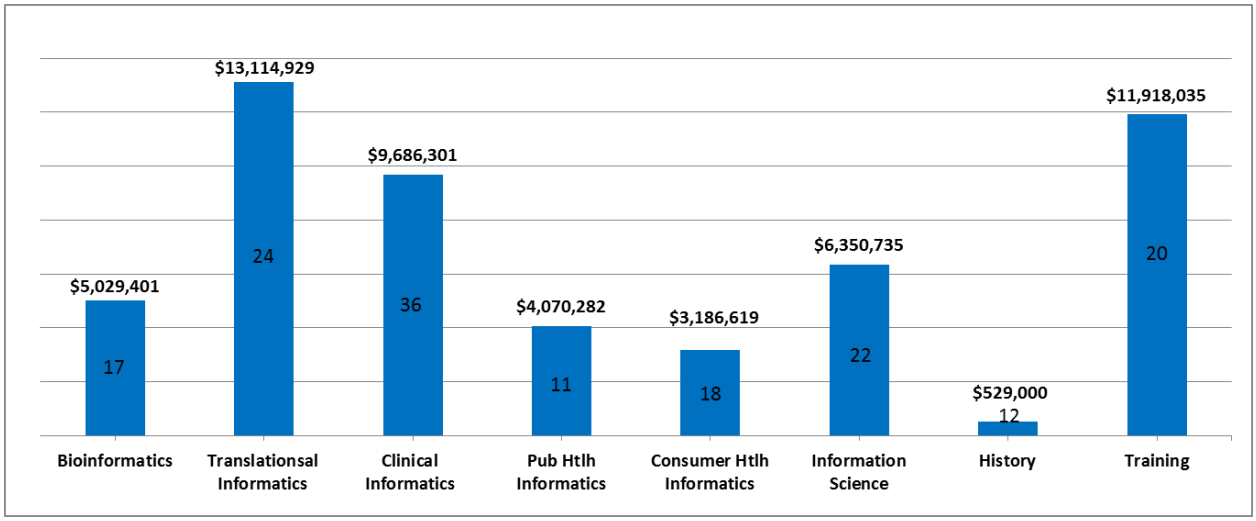
Figure 2 shows the distribution of NLM's grant funds by grant types. Appendix 3 provides a detailed breakdown by mechanism and activity.

Figure 2: FY2013 EP Grants Budget



These new awards bring NLM's active grant portfolio to 160 projects, valued at $53.9 million. Tables 6 depicts the distribution of these projects across NLM's investment areas by number and dollar value.

Table 6: **FY2013 Active Projects by Investment Area**



NLM issued a second funding announcement in its popular NLM Administrative Supplement for Informationist Services program. This unique collaborative funding program provides research grantees of other institutes with supplemental funds from NLM to add an in-context information specialist, usually a librarian, to their project team. Eight awards were issued in 2012. For the 2013 announcement, twelve other NIH Institutes joined NLM's announcement for this program, making their grantees eligible. The Association of Academic Health Sciences Libraries created a continuing education webinar series for its members featuring awardees of this program. Three webinars will be presented in October 2013.

NIH's Big Data to Knowledge (BD2K) initiative held several workshops during FY2013, focused on future research and training topics. Many of the attendees were NLM trainees or grantees. A funding announcement for BD2K Centers of Excellence was issued in FY2013 and will be reviewed in Spring 2014. Additional BD2K issuances will include calls for training for careers in big data, research and development of methods and tools for big data, one or more additional rounds of center grants, and support for important coordination tools, such as a data discovery index. NLM participates in all BD2K funding announcements.

Outcomes and Impact of NLM Research Grants

Current grantees published 325 articles acknowledging NLM grant support in 2013. Figure 3 shows the grant sources of those articles, which were published in 137 different journals.

Figure 3: **Publications from Active Grants in FY2013**

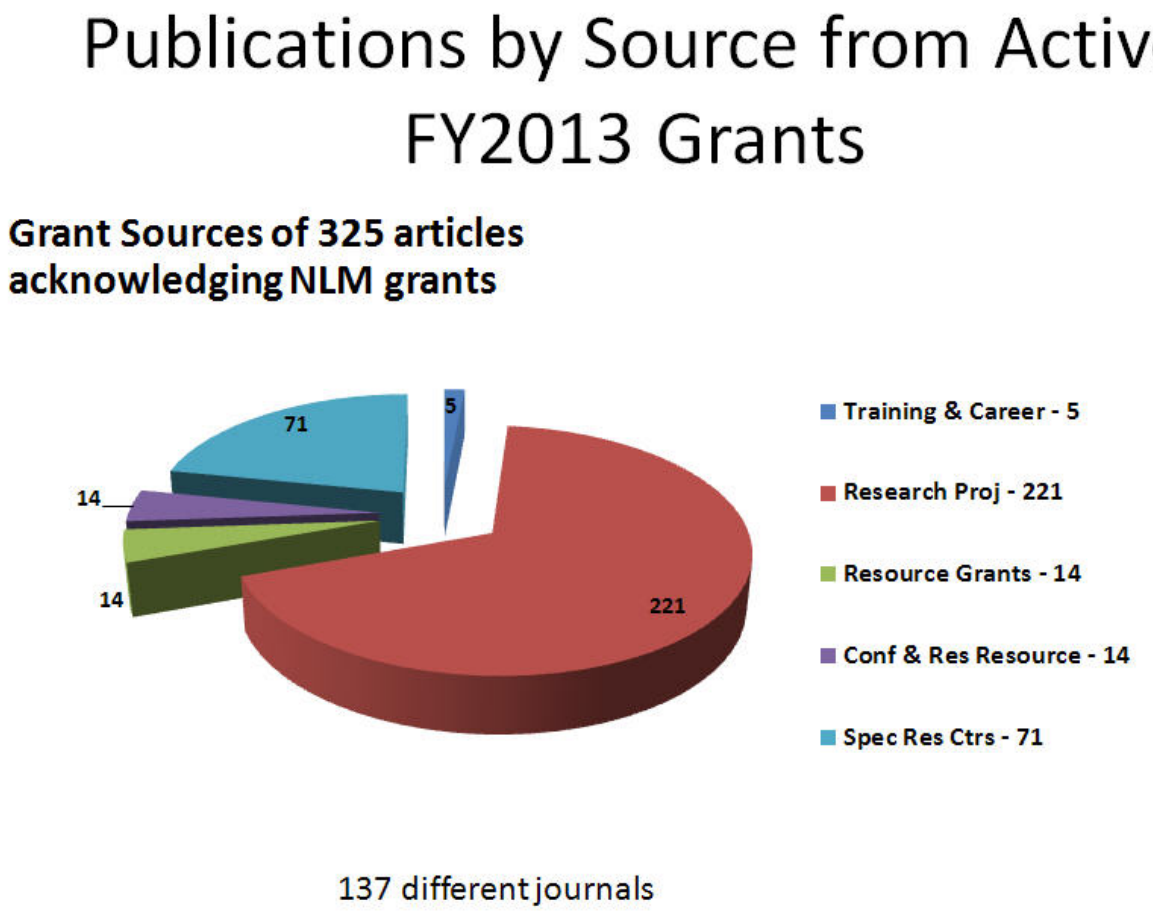
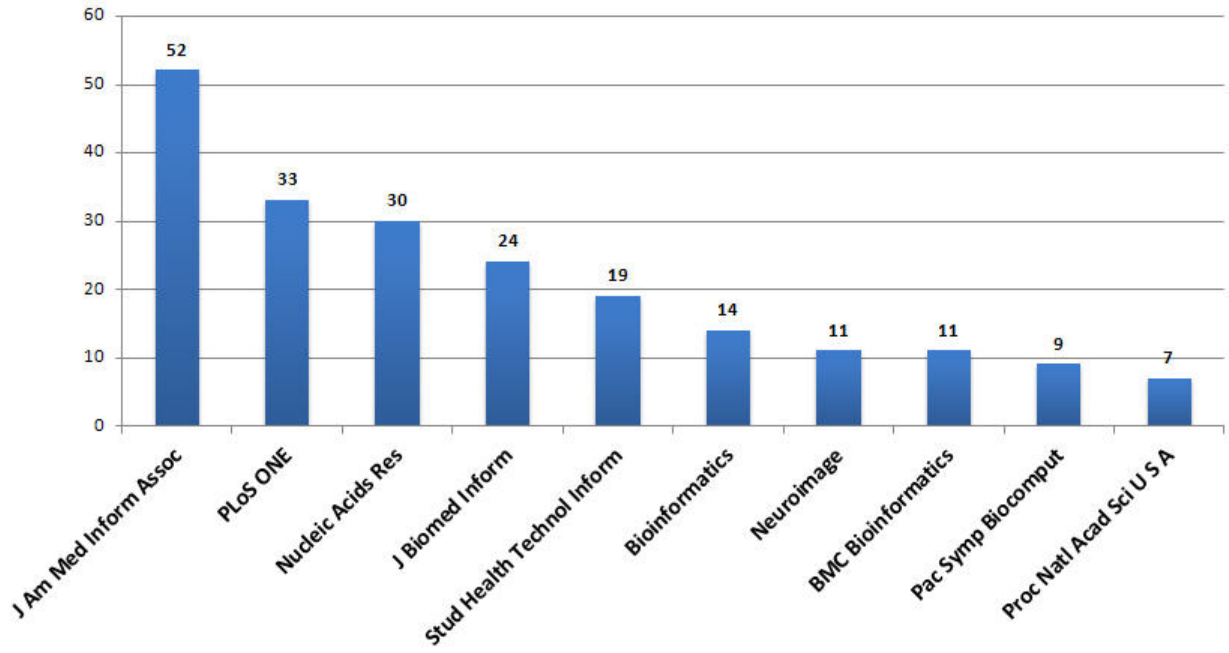


Figure 4 shows the top 10 journal titles in which these articles were published.

Figure 4: **Top 10 Publications where NLM Grantees Published, 2013**



Represents 210 of 630 journal articles published by NLM grantees in 2013, active and inactive grants

In terms of impact, SCOPUS lists 95 articles for 2013 and 258 articles for 2012 that acknowledge NLM grants. These 2012 and 2013 articles were published in 169 different journals.

Eighty-four percent of the 2012 articles have been cited at least once; 62 percent of the 2013 articles have been cited at least once, for a total of 2,311 citations. Over the past five years, 897 articles that acknowledge NLM grant support have been cited 14,847 times.

Honors for NLM Grantees in 2013

Dr. Atul Butte (Stanford) was honored by the White House with an Open Science Champion of Change award. Dr. Butte, who has authored more than 100 publications, is a former NLM trainee. His current NLM grant, "Integrating Microarray and Proteomic Data by Ontology-based Annotation", has produced 25 publications since 2011.

Dr. Peter Szolovits (MIT) received the Morris F. Collen Award for Excellence from AMIA. A leading researcher in the area of artificial intelligence for biomedicine, Dr. Szolovits's most recent NLM grant, "Capturing Patient-Provider Encounter through Text Speech and Dialogue Processing," ended in 2012. He leads one of the research components in "Informatics for Integrating Biology and the Bedside (i2b2)", an NIH Roadmap National Center for Biomedical Computing located at Harvard.

Dr. Nigam Shah (Stanford) received the 2013 AMIA New Investigator award. His NLM grant, "Methods for Generalized Ontology Terms Enrichment Analysis," was awarded in September 2013.

Dr. Joan Ash was senior author of the new SAFER Guides Series issued by the Office of the National Coordinator for Health IT. Her NLM grants on clinical decision support in electronic health records resulted in 25 publications between 2009 and 2013. Her *JAMIA* articles on unintended consequences of health care technology have been cited more than 1000 times.

Dr. Kwangsik Nho (Indiana) was awarded the silver medallion for his poster "Integration of

Bioinformatics and Imaging Informatics for Identifying Variants from Whole-exome Sequencing" at the 2013 AMIA Summit on Translational Bioinformatics. Dr. Nho is a former NLM trainee and current K99/R00 recipient.

Research Grants

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

Research Grant Program

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

Exploratory/Developmental Grants

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

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.

Small Business Innovation Research and Small Business Technology Transfer (SBIR/STTR) Research Grants

By law, all grant-issuing agencies set aside a portion of available research funds for Small Business Innovation Research (SBIR) grants. The 2011 reauthorization legislation instituted increases to the incremental percentage of the set-aside devoted to SBIR and to STTR for each of the next six years, beginning in FY2012. This year's set-asides were 2.7 percent for SBIR and 0.35 percent for STTR. NLM's SBIR/STTR interests for FY2013 focused on modeling tools for climate and environmental effects on human health, new technologies for integrating patient reported outcome data from electronic medical record (EMR) for development of clinical trials and tools to improve EMR usability. In FY2013, EP received applications through the NIH Omnibus Solicitation, and through a trans-NIH BISTI Funding Opportunity Announcement (FOA).

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 October 2012, an NLM Special Emphasis Panel reviewed 43 applications submitted for the Information Resource Grants to Reduce Health Disparities (RFA-LM-12-001) program. The focus of the RFA was on projects that bring useful, usable health information to health disparity populations and their health care providers through the use of computer and information science. Three new awards were made.

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 medicine and science. Three awards were made in this category in FY2013.

Training and Career Awards

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

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

Five-year institutional training grants support pre-doctoral, post-doctoral, and short-term informatics research trainees at 14 university-based programs across the country.

Collectively, the 14 continuing programs emphasize training in health care informatics (12

programs); translational bioinformatics (14 programs); clinical research informatics (12 programs); and public health informatics (9 programs). The National Institute of Dental and Craniofacial Research (NIDCR) also supports pre-and post-doctoral trainees in dental informatics at three of NLM's training programs: University of Pittsburgh, Oregon Health Sciences University and the University of Wisconsin-Madison.

In FY2013, NLM supported 113 predoctoral and 80 postdoctoral trainees at its training programs. The Short Term Trainee Slots (STTP) for Diversity program, which grew from 16 to 31 slots in 2012, was frozen in FY2013 due to sequestration and to a federal STEM consolidation initiative of the Office of Management and Budget.

Attendees at the NLM Informatics Training Conference, June 2013 in Salt Lake City, UT.

The annual NLM Informatics Training Conference was hosted by the University of Utah on June 18-19, 2013 in Salt Lake City, UT. Approximately 230 attendees including directors, faculty, staff, and trainees from all current NLM training programs; faculty and trainees from the Veterans Administration Informatics Training Sites; NLM staff; CDC staff; and guests. Research projects were presented in plenary and parallel sessions by 33 informatics trainees. There were 14 open-mic presentations during the conference which allowed trainees at early stages of their research to present their work in this popular brief format. An additional 27 trainees presented research-related posters at the meeting. Attendees voted for best speaker, best poster, and best open-mic talk. The award winners were:

- Best Open-Mic: Jacques Zaneveld, Rice University, "LRP6 May Cause a Novel Human Developmental Disease."
- Best Poster Day 1: Janet Woollen, Columbia University, "Engaging Hospitalized Patients with an Inpatient Personal Health Record."
- Best Poster Day 2: Meagan Whaley, Rice University, "Inferring Functional Human Language Pathways."
- Best Presentation Day 1: Jeremy Weiss, University of Wisconsin-Madison, "Multiplicative - Forest Continuous-Time Disease Prediction from EHRs."
- Best Presentation Day 2: Nicole Weiskopf, Columbia University, "Defining Completeness of Electronic Health Records for Secondary use: How Big is Your Database?"

Details of the agenda and a photo gallery are available at http://www.nlm.nih.gov/ep/trainingconf2013.html.

K99/R00 Pathway to Independence Awards

The NIH Pathway to Independence award (K99/R00) is a career transition program that combines a two-year mentored period (K99) with a three-year un-mentored research period (R00). Although applications to this program are not restricted to NLM's informatics trainees, they are the preferred applicants. To receive the R00 funds, K99 awardees must secure a career position with appropriate institutional support. Three new K99/R00 awards and three new R00 awards were made in FY2013.

K22 Independent Career Development Award

In FY2010, NLM re-launched its K22 early career award program. Several years of experience showed that a significant number of NLM's trainees, particularly those with MD degrees, were not applying for K99/R00 awards. This program was reinstated to meet their needs. Three new K22 awards were made in FY2013.

Early Recruitment to Informatics Research Careers

Summer Research Experience Program (R25)

Building on a similar successful American Recovery and Reinvestment Act (ARRA) program, NIH launched a new NIH Summer Research Experience Program in 2011, which provides a high quality research experience for high school and college students during the summer academic break. NLM limits participation in this program to its University-based informatics training programs (T15). Three awards were made in FY2011, to University of California Irvine, Vanderbilt University and Oregon Health Sciences University. In FY2013, these awards continued to support eight summer trainee slots at each program site.

Short Term Trainee Positions (STTP) to Enhance Diversity

Since 2007, NLM's Short-Term Trainee Program has focused on increasing diversity in the field of biomedical informatics by providing short term training experiences for under-represented minorities, and disabled or disadvantaged individuals. With the 2012 awards, this program grew from nine slots to 31 slots offered at 10 different sites. Typically, STTP trainees are appointed for three-month periods during the summer. However, the program was frozen in FY2013 due to sequestration and a

federal STEM consolidation initiative imposed by the Office of Management and Budget.

Trans-NIH Initiatives

Since October 2010, NLM has participated in the trans-NIH Basic Behavioral and Social Science Opportunity Network. (OppNet). OppNet was designed to fund extramural research and training activities that explain basic mechanisms and processes that influence individual and group health-related behaviors, and that have implications across the prevention and disease spectrum. The program, housed within the NIH Office of the Director, made its final awards in 2013. In 2013 NLM hosted one of the OppNet funding opportunity announcements, Basic Social and Behavioral Research on Culture, Health, and Well-Being (R24, RFA-LM-12-2002). This program provided grants for infrastructure support to develop, strengthen, and evaluate trans-disciplinary approaches and methods for basic behavioral and/or social research on the relationships among cultural practices/beliefs, health, and wellbeing. Ninety applications were accepted for review and seven were awarded using OppNet funds. NLM is the administrative home for one of these grants, awarded to Dr. Robin Gregory of Decision Research, Inc. for "Developing Indigenous Health Indicators to Reflect community Health Priorities."

Pan-NIH Projects and Interagency Collaborations

National Centers for Biomedical Computing (NCBC)

The National Centers for Biomedical Computing (NCBC) are cooperative agreement awards originally funded under the NIH Roadmap, but transitioned to Institute funds across the final five-year period. In 2013, the program ended its 10-year funding cycle with 100 percent of funds provided by the sponsoring Institutes. NLM continues to administer one NCBC center, "Informatics Integrating the Bench and Bedside (i2b2)," based at Harvard University's Brigham and Women's Hospital. NLM program officers have held scientific advisory roles in three NCBC centers: i2b2 at Harvard University, "NIH National Center for Physics-Based Simulation of Biological Structures (SIMBIOS)" at Stanford University and "integrating Data for Analysis, Anonymization and SHaring (iDASH)" at University of California San Diego. I2b2 is the first center grant NLM has funded, and is EP's most productive grant in terms of publications and software sharing.

Multi-institute Grant Programs

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

Applications for multi-institute programs are reviewed by the NIH Center for Scientific Review (CSR). Those that receive fundable priority scores are considered for awards alongside grants reviewed by NLM's study section. Links to the multi-institute initiatives in which EP participates are incorporated into the grant programs list on the EP Web site at http://www.nlm.nih.gov/ep/Grants.html.

Shared Funding for Research & Training Within NIH

In FY2013, NLM provided co-funding support to one NIH Director's Pioneer Award; an STTR award administered by NCI; an International Research Training Award issued by the Fogarty International Center (FIC), the iDASH NCBC administered by the National Heart Lung, and Blood Institute (NHLBI); and seven informationist supplements awarded to research grantees of other NIH Institutes. These collaborative funding arrangements represent co-funding from NLM to other Institutes in the amount of $490,099. In addition, NLM received $2,998,965 full or partial funding for eight NLM grants as follows: one OppNet grant, 2 NIH Director's Pioneer Awards, and one NIH Director's New Innovator Award from the office of the NIH Director; co-funding from the National Heart Lung and Blood Institute (NHLBI) for the I2B2 Center; co-funding from the National Institute of Dental and Craniofacial Research (NIDCR) for dental informatics trainees at three NLM training programs.

Interagency Agreements and Funding

NLM grantee Dr. John Brownstein received $20,000 in supplemental grant funding support from the Office for the Assistant Secretary for Preparedness and Response (ASPR), DHHS, for a special health–related initiative relating to his HealthMap platform and data sources, aiding development of a tool to provide epidemiologic/disease outbreak information for hurricane impacted areas, resulting in improved situational awareness and timely analysis of disease threats.

Along with several other NIH Institutes, NLM provides support to the Protein Sequence Databank (PDB) at Rutgers University. NLM provided $177,300 in FY2013. The PDB, which is administered by the National Science Foundation (NSF), is the single worldwide repository for the processing and distribution of 3-D biological macromolecular structure data.

NSF and seven NIH Institutes, including NLM, issued a joint funding solicitation for research projects entitled "Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)". Hundreds of applications were reviewed by NSF study sections; reviewers used both NSF and NIH

criteria for applications with a health or biomedical research focus. NLM chose two projects from this initiative to transfer to NIH for funding consideration in FY2013; one was funded to Dr. Samantha Kleinberg of the Stevens Institute of Technology, entitled: "Core Techniques and Technologies for Advancing Big Data Science and Engineering."

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

In FY2013, there were 72,234 visitors to the EP Web site, a slight 6 percent decrease over the previous year. The EP Web site continued to have a substantial interest from international visitors who represented almost 25 percent of total visits. The most frequently viewed pages were those describing the grant programs, EP homepage, information pages such as the NLM FY2013 Funding Plan and Grant Deadlines, Frequently Asked Questions, and a page of links to NIH-produced tutorials for preparing grant applications. The EP Web site was enhanced by redesigning the homepage and the master page templates in accordance with NLM design standards. The homepage was reorganized to reflect the three primary EP user interests – research funding, career development, and review. The EP news box on the EP homepage was redesigned to include standard formatting for information about EP News, Grantee Spotlights, and the EP Lecture Series. Searchable archives were also added for these news sections.

Scientific Review

Biomedical Library and Informatics Review Committee (BLIRC): NLM's standing review group, the Biomedical Library and Informatics Review Committee, evaluates grant applications assigned to NLM for possible funding. The BLIRC met three times in FY2013 and reviewed 116 applications (as compared to 152 in FY2012). The Committee (Appendix 4) reviews applications for biomedical informatics and bioinformatics research projects, knowledge management/applied informatics projects, career support awards, and fellowships.

Special Emphasis Panels (SEPs): Five Special Emphasis Panels were held during FY2013 compared to four in FY2012. These panels are convened on a one-time basis to review applications for which the BLIRC lacks appropriate expertise, such as Scholarly Works grant applications, when a direct conflict of interest exists between the application and a member of the BLIRC, or when the number of applications received is simply too large for BLIRC to handle. Overall, NLM's SEPs reviewed a total of 96 applications during FY2013, compared to 64 in FY2012.

Grant Review Activities: Overall, 340 applications were reviewed for which NLM was the primary assignment. Of those, 212 were reviewed by NLM. The remaining 128 applications were reviewed by CSR including SBIR/STTR and R24 applications. Of the applications reviewed by NLM, 51 percent were in one of the three research grant mechanisms (83 R01s, 25 R21s, and 1 R13). Career Development Awards (K99s and K22s) represented 10 percent (21 applications), Scholarly Works (G13s) represented 18 percent (39 applications) and Resource Grants (G08s) represented 21 percent (43 applications) of the total applications reviewed.

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 Requests for Applications (RFA). In FY2013, the subcommittee held two early concurrence panels; 21 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 2,410 NLM grant applications were included in the en bloc votes (340 primary and 2,070 dual).

Appendix 1: RFA/PA Actions in FY2013

NLM's Core Active Grant Programs

Announcement	Title	Expiration
PAR-13-300	NLM Express Research Grants in Biomedical Informatics (R01)	September 8, 2016
PA-13-302	Research Project Grant (NIH Parent R01)	September 8, 2016
PA-13-303	NIH Exploratory/Developmental Research Grant Program (Parent R21)	September 8, 2016
PA-13-347	NIH Support for Conferences and Scientific Meetings (Parent R13/U13)	September 8, 2016
PA-13-313	Academic Research Enhancement Award (AREA) (Parent R15)	September 8, 2016
PA-13-276	NLM Administrative Supplements for Informationist Services in NIH-funded Research Projects (Admin Supp)	November 6, 2013
PAR-13-014	NLM Grants for Scholarly Works in Biomedicine and Health (G13)	February 21, 2014
PAR-13-284	NLM Career Development Award in Biomedical Informatics (K01)	September 8, 2016
PA 11-197	NIH Pathway to Independence Award (K99/R00)	September 8, 2014

Multi-institute Active Announcements in Which NLM Participates

Announcement	Title	Expiration
PA-12-150	Research Supplement to Promote Reentry into Biomedical and Behavioral Research Careers (Admin Supp)	September 30, 2015
PAR-13-130-132	Understanding and Promoting Health Literacy Research Grants (R01) (R21)	May 8, 2016
PA-13-292-288	Behavioral and Social Science Research on Understanding and Reducing Health Disparities (R01) (R21)	September 8, 2016
RFA-LM-12-002 RFA-EY-13-001	NIH Basic Behavioral and Social Sciences Opportunity Network (OppNET) Grants	December 18, 2012 November 1, 2012
PA-13-234	Small Business Innovation Research (SBIR) Grant (R43/R44)	January 8, 2014
PA-13-235	Small Business Technology Transfer Grants (STTR) (R41/R42)	January 8, 2014
PAR-09-220/221	Innovations in Biomedical Computational Science and technology Initiative (SBIR [R43/R44]) (STTR [R41/R42])	September 8, 2013
PA-12-196	Innovative Health Information Technology for Broad Adoption by Healthcare Systems and Consumers (SBIR)(R44)	January 8, 2015

Announcement	Title	Expiration
PAS-10-226	Advancing Novel Social Science Research on Understanding and Reducing Health Disparities (R01) (R21)	January 8, 2013

Appendix 2: NLM New Grants Awarded in FY2013

NLM NEW AWARDS FY2013

Research Grants (R01)

COHEN, TREVOR (NEW INVESTIGATOR)
Using Biomedical Knowledge to Identify Plausible Signals for Pharmacovigilance
1 R01 LM011563-01A1
University of Texas Health Science Center Houston

DAVULURI, RAMANA V
Informatics Platform for Mammalian Gene Regulation at Isoform-level
1 R01 LM011297-01A1
Wistar Institute

DEL FIOL, GUILHERME
Meeting Clinicians Information Needs with Highly Tailored Knowledge Summaries
1 R01 LM011416-01
University of Utah

FORAN, DAVID J
Image Mining for Comparative Analysis of Expression Patterns in Tissue Microarray
2 R01 LM009239-05A1
Rutgers Biomedical and Health Sciences - Cancer Institute of New Jersey

FRIEDMAN, CAROL
Pharmacovigilance Methods: Leveraging Heterogeneous Adverse Drug Reaction Data
2 R01 LM010016-05
Columbia University Health Sciences

HRIPCSAK, GEORGE M
Discovering and Applying Knowledge in Clinical Databases
2 R01 LM006910-14
Columbia University Health Sciences

KLEINBERG, SAMANTHA
BIGDATA: Causal Inference in Large-Scale Time Series with Rare and Latent Events
1 R01 LM011826-01
Stevens Institute of Technology

MIKLER, ARMIN ROBERT
Minimizing Access Disparities in Bio Emergency Response Planning
1 R01 LM011647-01
University of North Texas

MOORE, JASON H
Bioinformatics Strategies for Genome-Wide Association Studies
2 R01 LM010098-05
Dartmouth College

SHAH, NIGAM
Methods for Generalized Ontology Terms Enrichment Analysis
1 R01 LM011369-01A1
Stanford University

TONELLATO, PETER J
Predictive Optimal Anticlotting Treatment for Segmented Patient Populations
1 R01 LM011566-01
Harvard University (Medical School)

WAGNER, MICHAEL MATTHEW
Probabilistic Disease Surveillance
1 R01 LM011370-01A1
University of Pittsburgh at Pittsburgh

Exploratory/Developmental Research (R21)

LASKO, THOMAS
Scalable Biomedical Pattern Recognition via Deep Learning
1 R21 LM011664-01
Vanderbilt University Medical Center

Applied Informatics Resource (G08)

FISCELLA, KEVIN
Addressing Disparities in Health Information through an FQHC-Library Partnership
1 G08 LM011524-01
University of Rochester

NUNLEE-BLAND, GAIL LOUISE
Living Smartly with Diabetes: Using PWP and Mobile PWP for Self-Management
1 G08 LM011545-01
Howard University

ZENG, QING
Graphics to Enhance Health Education Materials for Underrepresented Populations
1 G08 LM011546-01
University of Utah

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

KUIKEN, TODD ALAN
Preparation of a Monograph on Targeted Reinnervation
1 G13 LM011221-01A1
Rehabilitation Institute of Chicago

RUSNOCK, ANDREA A
The Birth of Vaccination
1 G13 LM011206-01A1
University of Rhode Island

SCHIEBINGER, LONDA
Race and Human Experimentation in the Eighteenth-Century Atlantic World
1 G13 LM011470-01
Stanford University

Career Transition Award (K22)

COOPER, LEE
Multiscale Framework for Molecular Heterogeneity Analysis
1 K22 LM011576-01
Emory University

EL-KAREH, ROBERT E
Minimizing Physician Errors: Feedback of Patient Outcomes after Handoffs
1 K22 LM011435-01
University of California

MCCOY, ALLISON B
InSPECt: Interactive Surveillance Portal for Evaluating Clinical support
1 K22 LM011430-01A1
Tulane University of Louisiana

NIH Pathway to Independence Award (K99)

CONWAY, MICHAEL AMBROSE
Utilizing Social Media as a Resource for Mental Health Surveillance
1 K99 LM011393-01A1
University of California

MAYERICH, DAVID
Large-Scale Reconstruction of Microvascular Networks and the Surrounding Cellular
1 K99 LM011390-01A1
University of Illinois Urbana-Champaign

NABAVI, SHEIDA
Novel Integrative Method to Detect Biomarkers of Breast Cancer Resistance
1 K99 LM011595-01
Harvard University (Medical School)

Small Business Innovations Research (SBIR) and Small Business Technology Transfer (STTR) Awards (R41, R42, R43, R44)

BOTKIN, MATT
A Search Engine for Heterogeneous Information Needs in the Clinical Workflow
1 R43 LM011590-01
Medsocket of Missouri, Inc.

SEGAL, MICHAEL M
Interoperable Decision Support to Improve Diagnostic Workflow across Multiple EHR
2 R44 LM011585-02
Simulconsult, Inc.

Appendix 3: FY2013 Grant Budget by NIH Mechanism and NIH Activity Code

FY2013 Operating Budget Request by NIH mechanism groupings	No.	FY2013 Budget Total Amount
Research Project Grants (R01, R21, R00,)	84	$23,040,712
SBIR/STTR (R41, R42, R43, R44)	2	$708,763
Other Research - Research Careers (K99, K22)	8	$931,830
Other Research - Other (G08, G13, R13, R25, D43)	26	$2,767,685
Training - Institutional (T15)	15	$10,490,071
R&D Contracts (Y03)	1	$177,300
National Centers for Biomedical Computing Award (U54)	1	$3,067,637
EP Grant budget excluding TAPS and Operations	**137**	**$41,183,998**

FY2013 Grant Expenditures by NIH activity code	No.	FY2013 Budget Total Amount
D43: International Training Grants in Epidemiology (cofund)	-	$227,600
DP1: Pioneer Award	2	$192,060
G08: Knowledge Management & Applied Informatics; Planning Grant for IAIMS	7	$642,048
G13: Scholarly Works in Biomedicine and Health	11	$459,907
K99: Pathway to Independence	4	$336,942
K22: NLM Independent Career Development Award for Biomedical Informatics	4	$594,888
	-	-
	-	-
	-	-
P41: Biomedical Resource Grant	2	$1,204,774
R00: Pathway to Independence	10	$1,695,299
R01: Research Project Grants	67	$20,349,107
	-	-
R13: Conference Grants	3	$65,352
R15: Academic Research Enhancement Award (AREA)	-	-
R21: Exploratory/Developmental Grants	5	$804,246
R25: Education Projects	3	$168,004
R41: Small Business Technology Transfer (STTR)	-	$81,334
R42: Small Business Technology Transfer (STTR)	-	-
R43: Small Business Innovation Research (SBIR)	2	$627,429
R44: Small Business Innovation Research (SBIR)	-	-
	-	-
T15: University Biomedical Informatics Research Training Programs	15	$10,490,071
	-	-

FY2013 Grant Expenditures by NIH activity code	No.	FY2013 Budget Total Amount
U54: NCBC Roadmap Center	1	$3,067,637
Y03: Inter-Agency Agreement	1	$177,300
EP budget excluding TAPS and Operations	**137**	**$ 41,183,998**

* includes 7 informationist supplement awards

OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

Ivor D'Souza
Director

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

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

Controlled Medical Vocabularies

Unified Medical Language System (UMLS)

This year we made substantial improvements to the systems used to edit and release new versions of UMLS Metathesaurus.

We migrated from a Solaris to a Linux platform, which resulted in the Metathesaurus Editing and Maintenance Environment (MEME) which radically decreased processing times. This improvement will shorten the UMLS release cycle by several days, providing more time for product testing prior to release. We have improved the source integrity formats during production, and the quality of production with continual updates of the UMLS live database as sources are inverted and inserted. We developed new tooling to invert and insert nine new terminology sources used by CMS for clinical quality measurement, all valuable additions to the UMLS Metathesaurus.

We provided new capabilities to improve access to the RxNorm prescribable subsets from within the UMLS and to facilitate the integration of SNOMED CT content within RxNorm.

To achieve time savings, we set up separate Linux servers for the creation of new surveys that support the UMLS Annual Report application. This decreased the time required to create the surveys from a full day to mere minutes.

We also implemented 42 new medical terminology related items that users can obtain via the Internet by querying the UMLS Application Programming Interface (API).

Value Set Authority Center (VSAC)

In coordination with Centers for Medicare and Medicaid Services (CMS) and Office of National Coordinator (for Health Information Technology (ONC), NLM implemented the Value Set Authority Center (VSAC)—an online portal that serves as the "source of truth" for terminology value sets used in in clinical quality measures required to achieve Meaningful Use Stage 2. Other collaborators included the Agency for Healthcare Research and Quality, the Mayo Clinic as well as a number of commercial entities supporting the CMS's development of eMeasures (Electronic Health Quality Measures). VSAC is a repository for storage, search, and retrieval, and provides an authoring environment for creating purpose-specific subsets of standard terminologies known as value sets. VSAC provides eligible hospitals (EH) and providers (EP) access to the (value) sets of medical codes required to achieve MU2. This initiative has improved the quality of the value sets and reduced reliance on data keyboarding by humans. Four updates to the terminology content were released in FY2013 in accordance with CMS regulation for MU2 Certification.

Among the new capabilities added to the VSAC during the year is the ability to modify existing value sets and author new value sets via the VSAC Web portal. The new process gives measure developers the opportunity to select medical codes directly from NLM's collection of vocabulary standards and undergo a thorough quality check review prior to publication of the updated value set. By increasing the quality and timeliness of new value sets, NLM is making a significant contribution to the meaningful use of EHR systems.

International Health Services Terminology Standards Development Organization (IHTSDO)

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

This year OCCS released the SNOMED CT International Request Submission (SIRS) system. This application, hosted by NLM, provides new capabilities by allowing the IHTSDO community to submit and track requests for new medical terminology content in SNOMED CT, to replace a legacy Seibel system available from the IHTSDO.NLM will import into SIRS all current

legacy requests from the Siebel-based system. The SIRS system is based on the US SNOMED CT Content Request Submission (USCRS) system, which is currently used by NLM to support the processing of US-only content requests. This deployment establishes a core baseline of software to allow NLM to provide customized versions of the content request system for other IHTSDO member countries if they so desire.

OCCS took the initiative to provide added capability by publishing the first US Edition of SNOMED CT. The US Edition combines the International Edition and US Extension of SNOMED CT to provide end users with a complete single US SNOMED vocabulary set. The US Edition is published in the new RF2 format, as well as the legacy Release Format version 1 (RF1) format for backwards compatibility.

RxNorm

RxNorm is produced by NLM and is a standardized nomenclature for clinical drugs. Improvements were introduced this year to simplify the use of and improve the quality of RxNorm:.

- 898 brand names were transformed into their widely accepted drug label name (e.g., Percocet 5/325 was changed to Percocet) thereby improving the ability to retrieve all versions of Percocet
- A new source, the WHO's Anatomical Therapeutic Classification (ATC), was added to RxNorm. The addition of ATC content provides RxNorm users the ability to find a wider range of classes of drugs, including classes used in European countries;
- RxNorm also added new content. This included the release of two updates to the prescribable subset, which provides additional data for cross referencing allergy interactions and adds the ability to flag a subset of drugs that are not prescribable. This initiative serves to provide a more accurate picture of prescribable drugs on the market. RxNorm is now able to provide a more accurate representation of drugs that are no longer marketed.

The ten new versions of the RxNorm Editing System released this year provided new features for editors to perform their tasks more quickly and more effectively manipulate RxNorm data.

Medical Subject Headings (MeSH) and Related Systems

MeSH is the National Library of Medicine's controlled vocabulary used for indexing articles for MEDLINE/ PubMed.

We continuously develop new tools and enhance our existing products and methods. This year we added a MeSH Editing System spell checker. This allows MeSH editors to see, in real time, misspelled words against a dictionary consisting of American English words and the SNOMED CT—US Edition vocabulary. Users can now have the system report on any spelling errors in MeSH provisional records, allowing MeSH editors to correct the errors as appropriate directly from the interface.

We also updated the editing tools to give MeSH editors real-time access to other major medical terminologies stored in the NLM Unified Medical Language System (UMLS), thereby streamlining editing workflow.

Consumer and Public Health

MedlinePlus and MedlinePlus Mobile

MedlinePlus provides health information for patients, families and health care providers. Some MedlinePlus content is also available from a mobile-optimized Web site.

This year, we re-engineered the MedlinePlus system to reduce time taken to make changes to health topics, a core content type. The original process was designed and implemented long ago, making it difficult to expand to include new functionalities. By optimizing the health topics process we made it easier to maintain and expand and also improved the user experience. Simple changes that used to take two days can now be completed in two hours. Complex changes, which used to take two to three weeks, can now be accomplished in four to five days.

We also improved customer-facing mobile services by converting video content on the MedlinePlus Web site to more universally support all desktop and mobile platforms. By supporting Hypertext Markup Language fifth revision (HTML5), along with Adobe Flash technology, we are now able to provide native support to Apple iPads and iPhone devices.

MedlinePlus Connect

MedlinePlus Connect is a free service that allows electronic health record (EHR) systems to link to relevant and authoritative health information from MedlinePlus. It offers an Application Programming Interface (API) to provide external entities access to NLM health information services. It can match the Uniform Resource Locator (URL) of MedlinePlus Topics pages to the International Classification of Diseases (9th edition, Clinical Modification (ICD-9-CM) or Systematized Nomenclature of Medicine—Clinical Terms (SNOMED) codes they supply.

To assist with maintaining and enhancing the mappings use by MedlinePlus Connect, we created a working prototype management system for mapping medical codes to MedlinePlus content (topic pages, encyclopedia articles, and more). Users will be able to view the mapping and save it to a local database. This initiative will improve scalability, efficiency of creating and maintaining mappings, version control, and reliability and security of mapping data. An enhanced version of the management system is being developed. We also automated the NULL hit reports, to reduce time,

streamline the process, reduce manual intervention, and include ICD9 and SNOMED mapping with MedlinePlus health topics. Content personnel now can easily identify the most commonly requested codes and create new mappings, provide information to the public as needed, and reduce maintenance labor. Automating the NULL hit reports will save NLM 0.2 FTE in contract labor annually.

DailyMed

DailyMed provides health information providers and the public with a standard, comprehensive, up-to-date look-up and download resource of medication content and labeling as found in FDA approved medication labels (package inserts).

This year we implemented pharmacological drug classes in DailyMed, including browsing by pharmacological class and downloading pharmacologic class indexing files. Since marketed drugs change constantly when new drugs are added or existing drugs are taken off the market, there is an advantage to reference drug classes in Health IT applications instead of individual drug names. This new feature will improve search results, and help health care professionals build more flexible systems for customized health records, e-prescribing, and related uses.

To help health care providers, patients, and insurance companies better use the billing unit standards for the processing of drug claims, we enabled downloads of a billing unit index file from DailyMed that should decrease the erroneous invoicing of rebate payments, and more.

We implemented new application programming interfaces (APIs) to provide public access to NLM systems. We implemented a set of APIs that allow the public to obtain FDA's structured product labeling (SPL), from the NLM DailyMed Web site, that match an RxNorm drug concept or an SPL Group version. This initiative supports improving patient safety, and is consistent with the President's Digital Government Strategy, which looks to make high-value Government data easily accessible by other agencies and the public.

NLM Digital Repository

The NLM Digital Repository project supports NLM's 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.

This year, OCCS designed, developed and deployed a completely new architecture for the NLM Digital Repository to improve response time, usability, and maintenance. This included an entirely new primary user interface, and upgrades to all hardware and software components. OCCS heavily customized and integrated the selected user interface, the open-source Ruby on Rails discovery tool Blacklight, into the Digital Repository system. All major software components, including Fedora, Solr, GSearch, and Tomcat were upgraded. New servers were configured and deployed for increased scalability and maintenance, and the new system was tested for accessibility, security, load, responsive design, usability and functionality. Millions of pages of content were migrated to the new system, and the underlying database and search indices were rebuilt and optimized. OCCS also designed and delivered a new Digital Repository Web service, using infrastructure of the Data Explorer tool.

We extended support for multiple collections. NLM currently has content from five different collections, growing beyond the original Medical Heritage Library collection. For the Medicine in the Americas 2 Project, OCCS built upon the existing Medical Heritage Library (MHL) collection digitization workflow and developed generic processes to support any collection digitized and processed using the Kirtas Book Scan Editor (BSE) equipment used for the MHL project. The MHL workflow was tuned for routine NLM use, and no longer receives funding assistance from the MHL. This new generic workflow was used to support digitization, preservation and presentation of new collections, including Shakespeare Humors, and NLM Publications.

Digitization Scan Quality Assurance System

OCCS re-developed a Web-based quality assurance (QA) system for the NLM Digitization Program to automate, improve and streamline QA processes, using automated scripts to check the quality of digital artifacts generated from the scanning of historical medical literature. We re-engineered the system to provide data processing, workflow, and analysis using various tools, as well as logging and tracking of each scanned image. This effort has enhanced the functionality and improved the performance of the system, further reducing manual effort.

As a result of the new design and automation, the time required per week by IT staff used to fix quality issues related to the scanning has dropped from eight hours to zero. The Digital Scan QA process has also reduced the manual labor required to validate the quality of digital artifacts, by 60 percent (equivalent to two FTEs).

Using Cognos to map the underlying data structure, we provided the NLM's Digital Repository curation team the ability to create their own reports that capture the progress of digitizing NLM's historical collections to digitized artifacts. Prior to this improvement, it took 16 hours to generate each new report, or a total of 800 hours for roughly 50 unique reports. In the new system, the curation team can generate their own reports in a quarter of the time—a reduction of 75 percent in process time. This also provided reliability and data and format consistency for all Digital Repository reports. Staff

can create multiple ad-hoc reports with this package and can also schedule reports on a daily basis, or as needed.

Health Services Research Projects in Progress (HSRProj)

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

The major initiative this year added the capability for the HRSProj Web site to display multiple project funders (Supporting Organizations) and annual funding amount for each year of the project. This allows users to track continual funding for a project and to see all funders contributing to the fiscal year award amount.

Responsive Web Design

Responsive Web Design (RWD) is a Web design approach which provides an optimal viewing experience, easy reading and navigation across a wide range of devices (from desktop computer monitors to mobile phones).

OCCS has taken the lead in redesigning the Health Services Research Information Central (HSRIC) Web site by applying responsive design principles. We provided a single Web site that serves the public regardless of the device of Web browser that a user utilizes to access HSRIC. This has improved the efficiency of development and maintenance.. There is only one Web site to design and maintain versus the separate desktop and mobile designs required before.,

Project Tracking System

To improve issue and project tracking for our application development teams, OCCS has implemented . Jira, a proprietary product developed by Atlassian, for bug tracking, issue tracking and project management.

Implementing Jira has improved the efficiency of project, task, bug, and request tracking; helped us to manage our resource and team workloads; and had a solid impact on our productivity and results. Jira has simplified every step for everyone involved. We can now document, review, and resolve issues with ease within one system, but Jira also integrates easily with Confluence (NLM Wiki tool) and other tools. Jira's features, such as workflow, can streamline processes with increased accuracy, consistency and timeliness.

IT Infrastructure Services—NLM High Speed Communication Network

NLM has deployed and maintains a state-of-the-art communications network to fulfill its mission of collecting, organizing, and disseminating biomedical information throughout the world. To meet ever-increasing demands for new network services and higher bandwidth, OCCS continually maintains and enhances the network infrastructure and related services. The network is designed to be reliable and secure in order to support interconnection of its internal computing resources as well as high speed connections to external networks, i.e. the NIH campus network, the NIH Consolidated Collocation Site (NCCS) in Sterling, VA, the public Internet, and the Internet 2 research network. These network facilities enable access to, and delivery of NLM information services for staff, contractors, partners, and users..

Network Capacity Improvements and Readiness

OCCS maintains a contract with Level3 in order to continue NLM's connection to the commodity Internet. The dedicated bandwidth of the NLM/Internet connection is 3 Gbps (Gigabits per second), with a capability to burst up to 10 Gbps as needed. NLM's connection to the Internet2 network, via the Mid Atlantic Exchange (MAX) at the University of Maryland, remains at two 10 Gbps links.

This year, work began toward increasing the bandwidth of the NLM/Internet2 connections from two 10 Gbps links to one 100 Gbps link with a 10 Gbps backup. A new Juniper router was procured and installed at the NLM perimeter. This router is capable of supporting the 100 Gbps connections required. The Internet2 connections support the high speed transfer of large data sets between NLM and research partners via the Internet2 network, as well as general biomedical information to NLM clients

We redesigned and updated sections of local area network (LAN) cabling distribution components within Buildings 38 and 38A. This project replaced old LAN access switches and increased uplink speed by 900 percent from 100 Mbps to 1 Gbps in selected areas of Building 38/B1. It also optimized LAN cabling by re-homing user connections based on floor assignments. The work enabled the removal of the outdated LAN Rack D from the 38/B2 area.

OCCS continued support for the Bethesda Hospitals' Emergency Preparedness Partnership (BHEPP) by participating in annual drills and maintaining network readiness.

Implementing Internet Protocol Version 6 (IPv6)

OCCS completed its successful introduction of Internet Protocol version 6 (IPv6), as an additional option to the older IPv4, available to access NLM's public-facing applications. We continue working on the next federal mandate, providing IPv6 access for internal systems to external Internet resources.

In order to meet the Office of Management and Budget (OMB) IPv6 mandate, OCCS introduced IPv6 running in parallel with the older IPv4 protocol. OCCS met the federal mandate to upgrade public/external facing

servers and services (e.g. Web, e-mail, Domain Name System, etc.) to operationally use native IPv6.

As a first step to meet the next IPv6 federal mandate, to have internal client applications that communicate with public Internet servers and support enterprise networks use native IPv6 by the end of FY2014, OCCS began testing and deploying IPv6 for an internal desktop pilot group. This work will continue in FY2014. Part of this effort is to select and deploy an IPAM (IP Address Management) solution for NLM. An IPAM system was procured in 2013, and will enable management of both IPv4 and IPv6 IP addresses. It will also provide for centralized (Dynamic Host Configuration Protocol (DHCP) and management of Domain Name System (DNS) services within NLM.

Unified Communications Evaluation Started

OCCS began implementing a pilot project to evaluate the potential advantages to NLM offered by features of Unified Communications (UC) solutions. UC has the promise of providing a convergence of previously disparate products that include Voice Over IP (VOIP), voicemail, instant messaging/chat, conferencing capabilities (video, audio, and Web), and application and desktop sharing. VOIP implementation alone would reduce or eliminate antiquated and expensive Plain Old Telephone Service (POTS) and Integrated Services Digital Network (ISDN) telephone networks, reducing NLM phone costs, and enabling new functionality that will increase employee productivity.

IT Infrastructure Services—IT Security Initiatives

Protecting NLM's digital assets and vital information in cyberspace, against increasingly sophisticated threats, is a necessity to accomplish our mission. NLM continued to implement new security projects with preventive, protective, and responsive measures against existing and new cyber-attacks that have the potential to impair NLM's ability to provide critical information services. NLM has strengthened its security posture and fully supports the Department of Homeland Security (DHS) Continuous Diagnostics and Mitigation (CDM) Program.

Integrate HHS CSIRC Security Monitoring and Alerting

NLM worked with HHS Computer Security Incident Response Center (CSIRC) and NIH Incident Response Team (IRT) and has successfully introduced a high-speed intrusion detection system, a network monitoring system, and leveraged the HHS-wide security information and event management system. As a result, NLM is able to work closely with HHS and NIH security resources to promptly receive security alerts and improve speed and effectiveness of incident response to better protect NLM information technology from ongoing cyber threats. This collaboration also contributes to achieving agency-level IT security goals for security event monitoring, detection, analysis, eradication, and recovery. These systems also support the DHS Continuous Diagnostics and Mitigation (CDM) Program.

Centralized Security Logs Facilities Security Event Correlation

To improve NLM's overall security posture, NLM acquired and introduced a system to centralize and correlate operational and security-related log files. This solution enables instant visibility and event correlation across logs collected across the network from various devices. This centralized log collection pulls log data tuned using system configuration templates, and indexes these for optimal search and analysis. This log collection expands our situational awareness, and through real-time event monitoring, alerting, and log analysis improves our ability to detect and manage incremental or distributed attacks, and speeds incident investigations. This capability directly supports the DHS Continuous Diagnostics and Monitoring (CDM) Program to continuously monitor and alert on abnormal events, and provide security event correlation and feeding of security event data to the HHS CSIRC.

Passive Vulnerability Scanner Improves Real-Time Situational Awareness

Following the successful deployment of a centralized vulnerability management system at NLM, another supporting capability of the DHS CDM Program, the NLM Security Team introduced a service that takes NLM's centralized vulnerability management system one step further by examining live network data feeds against known vulnerabilities. This correlation of live network data against security scan results produces an accurate alerting mechanism for hacking attempts on known system vulnerabilities. This "passive vulnerability scanner" is an additional layer of inspection that improves NLM's overall security posture, as it refines the accuracy of security alerts and further bolsters the situational awareness with real-time, meaningful events.

Server Compliance Scan for All NLM Divisions

Implementing the vulnerability management system also revealed an opportunity to more consistently and deeply inspect systems in order to accurately identify non-compliance and vulnerable applications. The NLM Security Team coordinated across all NLM Divisions to implement scans using authenticated (authorized) access credentials. This secure credentialed, authenticated scan allows near complete accuracy in identifying system software, open ports, and applications that are out of compliance. Throughout the year, NLM administrators worked diligently and remediated 1,633,741 reported vulnerabilities identified in the centralized vulnerability

management system by applying vendors' patches, tightening configurations, and implementing compensating controls. Across the NIH, NLM has been a leader in adopting credentialed scans, and stands out having among the highest vulnerability management success rates.

NLM Firewall Upgrade

OCCS has deployed various security monitoring and protective systems in order to guard its IT systems and services. These include preventive, protective, and detective controls, as well as response and compliance procedures. State-of-the-art firewalls shield the network at the perimeter and also at lower layers within the local network.

During FY2013, the NLM perimeter firewalls were upgraded from devices capable of processing 5 Gigabit per second (Gbps) of simultaneous input and output traffic to 10 Gbps bi-directional processing. These higher performing firewalls allow increased levels of network traffic to be supported through the NLM perimeter network, enabling NLM to meet increased bandwidth demands for mission critical applications. In addition, an upgrade of the internal OCCS-Public firewalls from the NetScreen to the Juniper SRX3600 platform increased firewall capacity for the OCCS public-facing network segments. This upgraded the underlying firewall technology and accommodates traffic growth.

HHS Trusted Internet Connection (TIC)

NLM participated in the planning for the HHS Trusted Internet Connection (TIC) network implementation that was scheduled to take place in FY2013; however TIC network implementation was delayed until FY2014. The Office of Management and Budget has mandated that all federal agencies reduce the number of connections to the Internet, and pass traffic through TIC approved security centers. HHS is providing its own TIC network, being implemented and operated via CIT/NIH networks and HHS security facilities. NLM obtained approval to pass "unrestricted" data traffic through its own established Internet and Internet 2 connections, and only "restricted" traffic needs to pass through the HHS TIC network. This will allow NLM to continue to meet its mission in the most efficient and cost effective manner with respect to communications and network services.

Implementation of Enterprise Incident Response and Forensics Solution

As part of NIH's effort to improve incident response and forensics capabilities, NLM deployed the Mandiant Intelligent Response (MIR) agent to all of NLM's Windows systems. The MIR agent performs complex inspection of each system to look for specific indicators of compromise (IOC), and to collect, on-demand, the necessary data for malicious activity analysis and forensics. IOC scans are conducted on a monthly basis automatically to identify potentially compromised systems based on indicators from NIH, HHS and other partnership agencies.

National Library of Medicine (NLM) Network Penetration Test

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

Desktop Security 2013 Cyber Climate

For desktop security, NLM's automated patch management program routinely applied over 165,000 patches on commodity desktops, fixing known vulnerabilities to software. More than 1,000 desktop computers were updated, with nearly 24 million signatures; that is a 20 percent increase from FY2012, demonstrating wide-spread viruses and continuous threats across cyberspace.

NLM's Web anti-virus service, a layer of protection against malicious code coming from Web traffic, has blocked over 7,176 security violations; that is a 26 percent increase from FY2012—an indication of an increased Internet security threat. This additional service layer protects against malware from the Internet and minimizes the burden on desktop security software alone to fend off threats to desktop computers.

IT Infrastructure Services—Green Computing and Energy-Saving Initiatives

Server Virtualization Reduces Power and Space

OCCS is using virtualization to rapidly deploy information systems, boost computing resource utilization, reduce server room footprint, and increase energy efficiency of NLM's mission critical applications. Virtualization is the foundation for our internal cloud strategy. This year OCCS increased the use of virtualization and green computing of server infrastructure by 12 percent. The number of virtualized servers has increased from 471 to 532 and number of physical servers reduced from 154 to 135. As of the end of FY13, 78 percent of all OCCS-supported servers are virtualized.

The chart below shows the additional progress in the Virtualization Program this fiscal year.

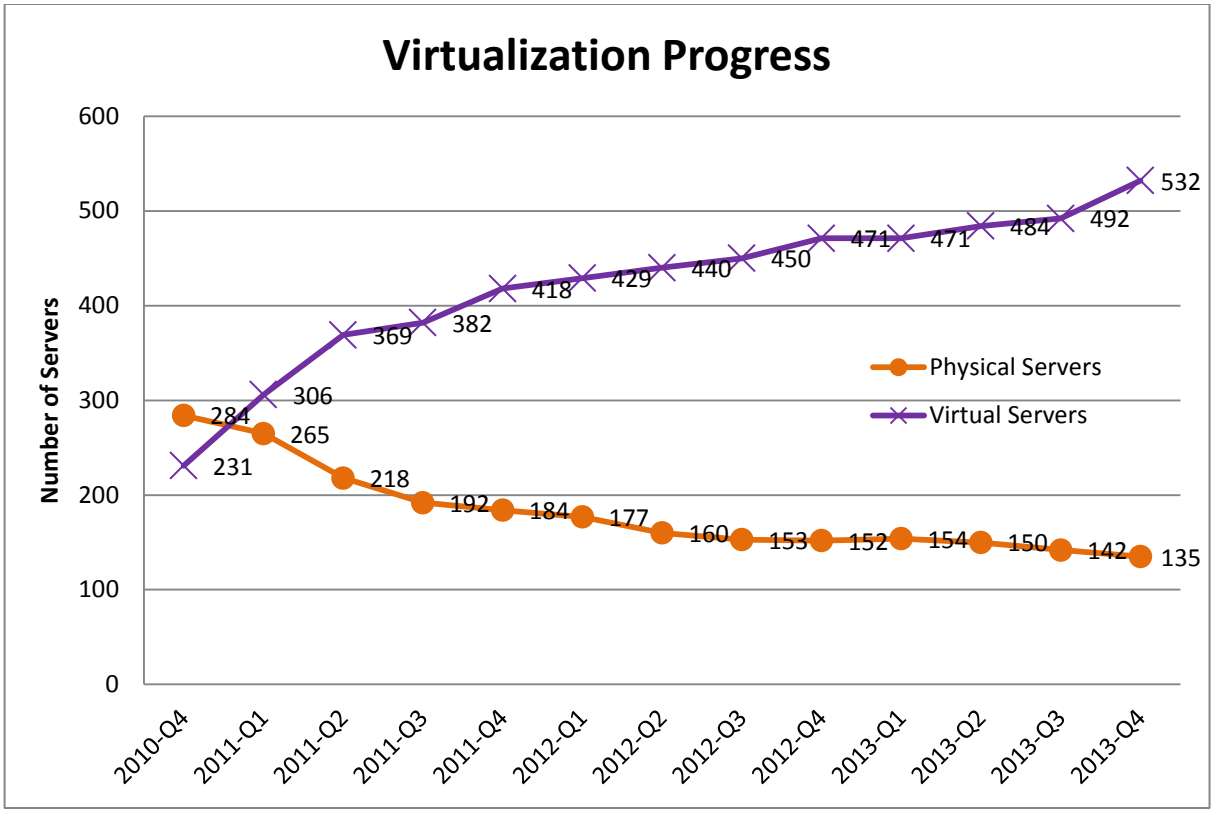


Figure 5: Virtualization Program Progress

During this same period, power consumption for these devices fell from 68 to 67 Kilowatts (kW), down from 110 kW at the start of FY2011. Rack space usage also declined during this year, falling from 696 to 661 rack units, a savings of about one half of a Data Center cabinet. (A rack unit "U" is equal to approximately 1.75" of the vertical face of a rack, where a typical rack can hold 42U.) Also since FY2011, the Virtualization Project collapsed our Data Center footprint from 1043U to 661U of cabinet space, extending our capacity for growth while reducing our energy consumption.

To support this transition to virtual servers, this year we also improved the capacity of our virtualization platform by upgrading obsolescing server hardware with a newer generation of servers. This increased available Central Processing Unit (CPU) by 25 percent and memory capacity by 35 percent at our NLM and NCCS sites.

Enterprise Computing and Storage Infrastructure Upgrade

This year we upgraded the OCCS enterprise storage systems (NetApp) residing at the NLM site, and increased network bandwidth of core servers and storage systems by tenfold, from 1GB to 10GB. This resulted in improved performance accessing large files, high-quality scanned images/video, and resiliency against peak demand. It also allows for future growth of data utilization. We continue to evaluate the storage infrastructure for opportunities to consolidate and ways to increase efficiency.

The chart below shows the growth of enterprise storage use over the past three years, measured in terabytes (TB).

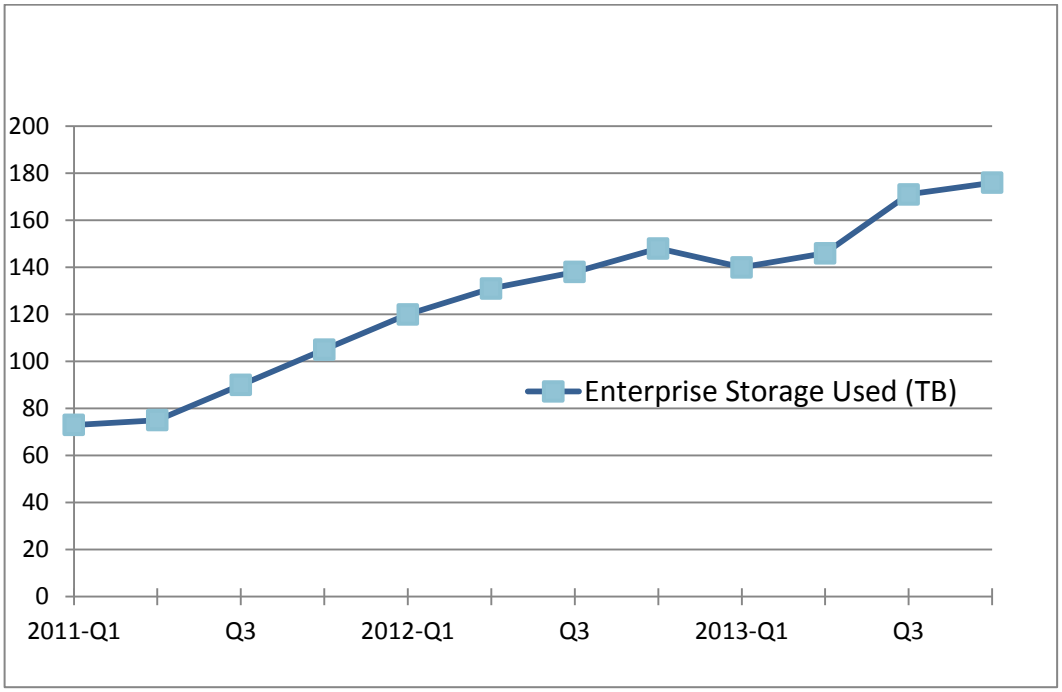


Figure 6: Growth of Enterprise Storage

NLM Data Center Redesign

In collaboration with other NLM teams, the OCCS's Facilities Management Section (FMS) conducted a redesign of the legacy NLM Data Center that will save over $200,000 a year on energy costs. FMS finished up its implementation of that multi-year program this year. Alternative power was added to the Data center from a second UPS system, which now provides for fully independent "A-B" power distribution capabilities, or N+1 redundancy. The total combined usable power is now at 1.6 MW with N+1 redundancy. Previously, the Data Center had been supported by four 500kW UPS systems in a parallel configuration with N+1 redundancy which allowed for 1.35MW of total usable power.

FMS has aligned the nine energy efficient Stulz CRAH cooling units installed last year, in accordance with the industry best practices, and to work efficiently within the NLM Data Center architecture. Coupled with separation of hot and cold air streams, cooling systems are operated more efficiently using higher chilled water supply and return temperatures. The higher chilled water supply/return temperature differential results in energy saving for pumping and smaller distribution systems. Facilitating this tuning, an Intuitive Control System (ICS) has been installed in the Data Center. ICS has proven its efficiency benefits in the high density zone data center, providing energy savings by automatically adjusting the CRAH cooling speeds based on the heat load at the rack level. The CRAH units' power supply was also distributed to improve redundancy and avoid a single point of failure.

This year, FMS also implemented a comprehensive leak detection system for the In-Row Coolers' overhead chilled water piping and CRAH units subfloor. In case of a water leak, the system instantaneously notifies NOSC personnel via the ISX Central, as to the exact location of a potential leak. This system speeds up NOSC personnel response time to a leak which if undetected, could cause extensive damage to IT equipment or harm to personnel.

Another environmental monitoring system was installed in the Onsite Alternate Colocation Facility (OACF). Multiple AC system failures were causing server issues due to unplanned shutdowns and switch problems. The OACF (located in Building 38) received an environmental monitoring system and was integrated with other NLM Data Center equipment monitoring systems to ensure efficient 24/7 oversight.

FMS replaced empty cabinet's space with a series of in row filler barriers to resolve stray mixing of hot and cold air within Data Center. This further ensured that the hot and cold aisles remain separated and minimizes mixing of hot and cold air. Staff also installed covers on the UPS battery terminals, and covers over a exposed sections of the power busway, to prevent accidental injury.

IT Infrastructure Services—Desktop and Mobile Infrastructure Initiatives

Increased Savings and Improved Management of Mobile Devices

The Desktop Services Order Fulfillment team continued to streamline NLM's wireless accounts with various carriers in order to comply with the President's Executive Order 13589. The team reduced the number of wireless lines by cancelling unused accounts and by implementing a wireless loaner pool program. The loaner program allows organizations to request wireless loaner equipment for a short time period instead of making a permanent purchase. As these devices can be provisioned from the loaner pool, more responsive provisioning is also offered. The wireless accounts are now categorized by UDL (User Defined Label) to allow for chargeback to respective departments. Role-based access to carrier's Web portals are positioned for account management.

Windows PC Security Operations Improved

To maintain and improve a high standard of Windows PC security operations, IBM Endpoint Manager (IEM) (formerly Big Fix) was tested and introduced this year. For a while, the IEM and the previous Shavlik server were run in service overlapped. After monitoring IEM for two months, the Shavlik product was decommissioned. IEM provides additional flexibility in how patches are distributed, and also provides greater accountability in ensuring all clients are patched, including offline PCs, due to its agent-based nature. IEM proved also to be faster overall in time spent scanning PCs.

Support for Digital Signage and NLM Video Wall

Successful deployment of the first two monitors of the Scala Digital Signage system took place in Building 38A. One monitor was set up in the NLM cafeteria along with a monitor run by NIH, and the other was set up in the lobby of the 38A/B1 level. The Scala server is housed in the B1 Data Center. The Scala presentation on these monitors contain content on announcements of NLM events, tour announcements, exhibit locations/dates, NIH events, staff promotions, and other announcements produced by the NLM Office of Communications and Public Liaison. Additional locations are being planned.

Nine monitors and a PC support an NLM video wall were also introduced in the B1 lobby area. The configuration of the video wall is now completed, and is displaying NLM Outreach Program content. This new device enables NLM to extend its outreach efforts beyond Native American pow-wows and conferences to also enlighten NLM staff and guests in the workplace.

IT Infrastructure Services—Web Tools

Web Analytics

NLM uses the WebTrends software package to track the number of pages served over time by the sites being managed and to provide detailed analysis of trends in site usage, audience composition, and other matters.

OCCS is working to improve Web usage analysis. This year we implemented the WebTrends Page Tagging capability. This has improved the accuracy of existing reports, as well as improved the efficiency of Web Trends servers, since page tagging requires less processing than processing Web logs. We have simplified the processing, and no longer need to collect Web logs from every Web server for every Web site analyzed by WebTrends.

Search Engine

Search facilitates users (both the public and NLM employees) in finding information that is available on the Internet or Intranet. This year we converted remote indexes to synchronized remote indexes. This has greatly increased the reliability of index updates since they are no longer dependent on network connectivity to all servers.

Business Intelligence and Report Writing Training

Cognos is IBM's business intelligence (BI) and performance management software suite. We use this software to enable business users to extract corporate data, analyze it and assemble reports. OCCS organized and provided training sessions to 18 Library Operations (LO) division staff members, on Cognos 10, saving more than $50,000 in training costs. In addition, we addressed issues/questions specifically relevant to NLM's business and concerns.

Database Platform Migration

We migrated standalone databases to a HP/Linux platform, which has increased the performance for application database operations, and has provided administrative efficiencies.

Medical Literature Support & Document Delivery Services

Data Creation and Maintenance System (DCMS)

OCCS achieved several performance improvements by updating the technology used for maintaining citations in MEDLINE, NLM's repository for journal citations and abstracts for biomedical literature.

The process time for loading publisher XML citations and readying them for data review and indexing has been reduced by 75 percent, going from one hour to 15 minutes each day. The process time for extracting data reviewed

and indexed citations for publishing has reduced by 66 percent, going from 1.5 hours to 30 minutes each day. The Year-End-Processing (YEP) process of extracting all citations for publishing has similarly been reduced threefold. Implementing this solution reduced the time needed for data exports during the baseline extraction, and it will simplify the data preparation side of each YEP cycle.

The remaining 12,000 OLDMEDLINE citations were processed, quality reviewed, then exported into DCMS. Over 2 million records pre-1964 records have been processed to date, greatly improving access to medical literature from the 1940s and 1950s.

DOCLINE

DOCLINE® is the National Library of Medicine's automated interlibrary loan (ILL) request routing and referral system. The purpose of the system is to provide efficient document delivery service among libraries in the National Network of Libraries of Medicine (NN/LM).

In an effort to reduce cost and save time, we rewrote the DOCLINE network quarterly reports using ColdFusion and Oracle queries. The new reports include are immediately available, easier to maintain, and do not require server space to store. This improvement resulted in the reduction of IT staff time by 0.5 FTE and customer/stakeholder time by 0.1 FTE.

Voyager Integrated Library System (ILS)

Voyager is a client/server protocol for fielded search and retrieval of MARC bibliographic records from the NLM integrated library system (ILS). Users can retrieve records in MARC-8 or UTF-8 character encodings.

The OCCS Voyager team automated the process of updating 380,000 authority and 25,000 bibliographic records in NLM's Integrated Library System using enhanced authority record information received from the Library of Congress. This improved the quality of NLM's authority and bibliographic records.

Internet Holdings and URL Maintenance (IHUM) Project

Library Operations' Technical Services Division (TSD) plans to remove holdings data in Voyager for serial titles with electronic access. In support of this initiative, OCCS made changes to the SEF (Serials Extract File) and SEF user applications, such as Serials Viewer, Manage Serials form, and the Cognos Package DCMS-LSTRC. Changes were also made to NLM Call Number conversion hierarchy and PubMed Central (PMC) attribute algorithm.

Outreach and Customer Services

OCCS staff continued to support the NLM Outreach Program through volunteering and providing technical support to over 30 outreach events, some of which are described briefly below. In addition, OCCS provided active support for several NLM exhibition initiatives.

Consumer Health Application (Outreach Project Database (OPD)) & Web Exhibits

The outreach projects database (OPD) is used internally to keep track of information such as who are the direct beneficiaries what kind of populations are served, funding, start and end dates, the planned evaluation and description of the projects.

We redesigned the Consumer Health Application (Outreach Project Database) and Web Exhibits. This redesign effort has reduced the amount of redundant data and time for data entry and replication. It has improved the maintenance and users' experience. During the redesign, new reporting capabilities were implemented, and a mobile version was created.

Online Contract Report

The main purpose of the Online Contract Reports is to digitize the annual and quarterly reports submitted by Regional Medical Library (RML) users, Center users and Subcontractors.

We added search features, advanced data analysis function for users, and integrated the contract report with other outreach applications under Single-Sign-On authentication. These changes have streamlined workflow for NLM and Office of Acquisition management staff who review and track work from the eight RML, three centers, and from RML Subcontractors.

Loading Dock Management System

We created the Loading Dock Management System for the Office of Administrative and Management Analysis Services (OAMAS). OAMAS was using a manual process to capture information about deliveries, which made searching for missing property very difficult. The implementation of the new Loading Dock Management System has helped speed up processing, and has improved manpower utilization. Overall, it has helped to efficiently manage government property and avoid the costs associated with missing property.

The loading dock system has helped property management track 529 accountable property shipments received, and 2,294 accountable assets decaled. It has helped the Property Management Team achieve reconciliation of accountable equipment reported in the "NIH Property Management Branch (PMB) Overages, Shortages and Un-decaled" reports.

Network Operations and Security Center (NOSC) Public Display System (NPDS) Upgrade

The NOSC display system consists of four 32-inch wide-screen plasma displays that are positioned in the NLM

Data Center windows. The intended audience of this display system is the general public and NLM staff. The system consists of information "panels" with descriptive text, statistical charts and near real-time activity monitors. Each panel focuses on a particular NLM service or IT infrastructure component. The panels include near-real-time utilization counters for MedlinePlus and for MEDLINE/PubMed, and NLM services as seen by remote users around the world. Near real-time utilization data for NLM's Internet-1 and Internet-2 data communications links are also displayed.

OCCS implemented new capabilities to upgrade the NPDS. This upgrade added new population-specific health mappings display to meet requirements from the Office of Communications and Public Liaison (OCPL), in line with the NLM's strategic plan objective of developing population-specific Web sites that focus on the issues of particular populations or geographic areas.

ADMINISTRATION

Todd D. Danielson
Associate Director for Administrative Management

Table 7. Financial Resources and Allocations, FY2013
(Dollars in Thousands)

Budget Allocation:

Program Area	Amount
Extramural Programs	$58,673
Intramural Programs	$246,121
Library Operations	$(71,741)
Computer and Communications Systems	$(28,832)
Lister Hill National Center for Biomedical Communications	$(44,055)
National Center for Biotechnology Information	$(84,089)
Specialized Information Services	$(17,404)
Research Management and Support	$13,627
Total Appropriation	$318,421
Plus: Reimbursements	$52,823
Total Resources	**$371,244**

Personnel

Table 8. FY2013 Full-Time Equivalents (Actual)

Program Area	Count
Office of the Director	8
Office of Health Information Programs Development	6
Office of Communication and Public Liaison	11
Office of Administration	55
Office of Computer and Communications Systems	47
Extramural Programs	21
Lister Hill National Center for Biomedical Communications	67
National Center for Biotechnology Information	281
Specialized Information Services	44
Library Operations	259
TOTAL FTEs	**799**

New Appointements

In December 2012, **Joyce Backus** was appointed the National Library of Medicine's Associate Director for Library Operations and a member of the Senior Executive Service. Ms. Backus has been at the NLM since 1985 when she joined us as an Associate in the NLM Associate Fellows Program. Since that time she has served in a number of key positions of increasing importance and responsibility, culminating most recently with her appointment as the Deputy Associate Director of Library Operations in June 2011. She has been the Acting Associate Director for Library Operations since June 2012 with the retirement of Sheldon Kotzin. With every new challenge she demonstrated extraordinary qualities of leadership, analytical ability, and management skill that have contributed immeasurably to the successes of NLM and Library Operations. Among many other achievements, Joyce was a key player in the development of MedlinePlus and led the creation and implementation of MedlinePlus Connect. She has been a highly effective representative of NLM and NIH in many contexts and in collaborations with other federal agencies, libraries, and professional associations. Joyce holds a Master's of Science Degree in Library Science from Catholic University of America and a Bachelor of Arts Degree in Sociology and English from Duke University.

In December 2012, **Mr. Daniel Hartinger** was appointed Director of the NLM Office of Acquisitions and Consolidated Operations Acquisitions Center. Mr. Hartinger joined us from the National Institute of Allergy and Infectious Diseases (NIAID), where he was the Chief of the Acquisitions Management and Operations Branch, NIAID Consolidated Operations Acquisitions Center. In that capacity, Mr. Hartinger managed a large branch responsible for major contracts, station support, research and development, information technology, and simplified acquisitions supporting the science of the NIAID intramural, clinical, and vaccine research programs, as well as the NIAID Office of the Director. Mr. Hartinger possesses a strong acquisitions background with 20 years of experience in the acquisitions field. He has held positions in a number of ICs, including NLM, and another federal agency, Coast Guard, through which he has gained both a broad perspective of federal procurement as well as the challenges facing acquisitions at the NIH. Mr. Hartinger holds a Bachelor's Degree in Political Science from the University of Delaware, a Master's Degree in Procurement/Contract Administration and Master of Business Administration from the University of Maryland University College.

In December 2012, **Mr. Duc Nguyen** was appointed as Chief, Medical Language Branch (MLB) in the Office of Computer and Communications Systems (OCCS). Mr. Nguyen will be responsible for engineering the technology platforms that support many of NLM's flagship medical terminology related applications such as the UMLS, RxNorm, MeSH, etc. Mr. Nguyen has over 30 years of experience in increasingly responsible roles within the information technology field, with specializations in software engineering and senior level program management, with various companies. In his most recent position, he was the Lead Computer Scientist within the Medical Language Branch. Previously, he worked in the private sector as a senior IT consultant, which included a brief stint at running his own IT consulting firm. Mr. Nguyen earned a Bachelor of Science degree in computer system engineering from

Western Michigan University, and two Master's Degrees in computer and communications sciences and in computer, Information and control engineering, from the University of Michigan. He is a certified Project Management Professional and the holder of two US patents.

In March 2013, **John Kilbourne, MD,** was appointed Head of the Medical Subject Headings Section (MeSH), Library Operations, having served as a Medical Officer within the MeSH Section since 2005. Dr. Kilbourne previously served as the lead on RxNorm, the NLM-developed standard terminology for clinical drugs. As the RxNorm lead, he established key relationships with industry and improved RxNorm data structures and data content to better serve users' needs. Additionally, he worked closely with other Federal agencies such as FDA, VA and CMS to facilitate NLM's role of providing drug terminology for use in e-prescribing, electronic health records, and FDA-approved drug labels. Dr. Kilbourne completed his residency and Board Certification in Family Practice and holds a Doctor of Medicine degree from the University of Illinois College of Medicine, Chicago. His prior experience includes terminology development for the College of American Pathologists, the creator of SNOMED CT.

In April 2013, **Dianne Babski** was appointed as the Deputy Associate Director for Library Operations (LO). She had been the Head of the MEDLARS Management Section in the Bibliographic Services Division in LO since 2008, having joined MMS as a librarian in 2005. As Head of MMS, she led NLM's 2011 "Show off Your Apps" challenge which showcased applications that make use of NLM data. She also chaired the NLM API committee, resulting in improved coordination and communication of these key tools for external developers. Ms. Babski also co-chaired the Library Operations strategic planning group on the Workforce of the Future which recommended many effective ways to better recruit and engage the LO staff in their professional development. Prior to joining NLM, she worked for the National Institute of Allergy and Infectious Diseases, NASA, the Henry M. Jackson Foundation, the Uniformed Services University of the Health Sciences, and the Walter Reed Army Institute. She has a Master's in information management and a BS in biology from the University of Maryland, College Park.

In June 2013, **John Kimbrough, MD, PhD**, joined the Office of the Director at the Lister Hill National Center for Biomedical Communications (LHNCBC). Previously, he was a staff cardiologist at the Orlando Veterans Affairs Medical Center in Florida. He received his BA in chemistry from Wesleyan University in 1991, PhD in physiology from the University of Rochester in 2000, and MD from the University of Rochester in 2000. He completed his internal medicine and cardiology training at Vanderbilt University. He is working on projects analyzing the MIMIC-II database to answer questions related to his interests in cardiology, internal medicine and other areas. His mentor is Clement McDonald, MD, Director, LHNCBC.

In July 2013, **Deirdre Clarkin** became the Deputy Chief, Public Services Division (PSD) in the Division of Library Operations, NLM. She had been the Head of the Collection Access Section, Public Services Division since December 2009 and prior to that had been Head of that Section's Onsite Unit since 2002. Most recently she has led the LO strategic planning initiative to study the future of resource sharing in the National Network of Libraries of Medicine. Ms. Clarkin serves as an NLM copyright expert and has taken multiple courses to further her knowledge in this complicated arena. Prior to her work at NLM, Ms. Clarkin held the position of Senior Librarian at the Federal Deposit Insurance Corporation. She has held reference and other public services positions in academic libraries and also a major law firm library. She holds a Master's Degree in library science, and is pursuing a Master's Degree in business administration, both from the University of Maryland in College Park.

NLM Associate Fellows Program for 2013 – 2014

Don Jason received his MLIS and MS in information architecture knowledge management with an emphasis in health informatics from Kent State University in 2013. Most recently, Mr. Jason worked as a library student assistant in the Oliver Ocasek Regional Medical Information Center where he specialized in reference, circulation and community outreach. He also worked as a graduate assistant in the College of Pharmacy at Northeast Ohio Medical University. In the beginning of his graduate career, Mr. Jason worked in several different capacities at Kent State University. He served as a diversity intern for the University Libraries, a graduate assistant for the School of Library and Information Science and as an instructional designer for the College of Communication and Information. During his graduate program, Mr. Jason successfully completed two Association of Research Libraries diversity programs, the Career Enhancement Program and the Initiative to Recruit a Diverse Workforce. He is also a proud recipient of an American Library Association Spectrum Scholarship. Mr. Jason received his BS in journalism from the E.W. Scripps School of Journalism at Ohio University.

Nicole Lehotsky received her MSLS degree in 2013 from the University of North Carolina - Chapel Hill. While completing her degree, she worked at UNC's Health Sciences Library as the NC Health Info Research Assistant, where she managed a team of volunteer editors who contributed information to the consumer health website and contributed ideas for a redesign of that same website to better meet the needs of residents in North

Carolina. Additionally, she interned in the Music Library at NPR in Washington, DC, and prior to that, worked as the Youth Services Humanities Coordinator at the Durham Public Library. She served as a founding co-chair of the Coalition of Youth Librarians, an organization dedicated to meeting the needs of library students wanting to work with youth and children, and received the Robert F. Asleson grant to attend the ALA Midwinter 2013 meeting. Prior to completing her graduate studies, Ms. Lehotsky worked as project manager and implementer for a healthcare IT company. She earned her undergraduate degrees in music education, music history, and English from Youngstown State University.

Katherine (Kate) Masterton received her MLIS from Drexel University in 2013. While obtaining her master's, she worked as a Library Services Assistant at the Drexel University Hahnemann Health Sciences Library providing support in access services, reference, and interlibrary loan. In addition, Ms. Masterton served as President of Drexel University's student chapter of the Special Libraries Association and volunteered at the Philadelphia FIGHT AIDS Library, which provides HIV/AIDS-related outreach and information services. She received the 2013 Beatrice Davis Education Award from the Philadelphia Regional Chapter of the MLA and a 2013 Knower Scholarship from the Philadelphia Chapter of the SLA. Prior to her time at Drexel, Kate served as an Americorps VISTA volunteer, worked and volunteered in academic, public, and state libraries, and received a BA in English from Hope College in Holland, MI.

Christian Minter received her MS in library and information science from The Catholic University of America (CUA) in Washington, DC in 2013. Ms. Minter was a recipient of the American Library Association Spectrum Scholarship and the Association of Research Libraries Diversity Scholarship. She served as the Vice President of the CUA Student Chapter of the Special Libraries Association. While completing her degree, Ms. Minter was employed as a library associate for the Prince George's County Memorial Library System where she provided reference and instruction services. During her program, she also completed a practicum at the Health Sciences and Human Services Library at the University of Maryland, Baltimore, and volunteered as an intern at the National Resource Center on ADHD in Landover, MD. Ms. Minter received her BA in biblical studies from Washington Bible College in Lanham, MD.

Holly Thompson received her MS in library and information science from the Palmer School of Library Science, at Long Island University in 2011. While completing her graduate degree, Ms. Thompson worked at the NYU Health Sciences Libraries and the Memorial Sloan Kettering Cancer Center library providing document delivery, circulation, and reference support. Ms. Thompson's more recent work as the inaugural NYU Health Sciences Librarian Fellow focused on the creation of online learning modules, LibGuides, and instructional support materials. Ms. Thompson also participated in the development of the NYU liaison program as a member of the NYU Health Sciences Libraries' division of Research, Education, and Clinical Support. As an undergraduate, Ms. Thompson worked as a library student assistant at the University of California, San Diego (UCSD) BioMedical Library where she first gained exposure to and interest in health sciences librarianship. Ms. Thompson received her BA in literature from UCSD.

Retirements

In October 2012, **Stuart J. Nelson, MD** has retired from Federal service after 16 years as head of the MeSH Section and 18 years in government. Under his leadership, the MeSH Section expanded beyond its name to include many services used by the medical informatics community, researchers, industry, health care organizations, and patients in the US and around the world. While at NLM, Dr. Nelson led the development of significant enhancements to the structure and content of MeSH, development of the MeSH translation maintenance system used in countries around the world, and improvements to content of the UMLS Metathesaurus. He was a key figure in the development of DailyMed as a dissemination vehicle for electronic structured product labels for medications submitted to the FDA and led the creation of the RxNorm clinical drug vocabulary, now a US national standard for electronic health records. He received the NLM Regents Award for this latter achievement. While at NLM, Dr. Nelson was instrumental in significant collaborations between NLM, the FDA, and the Department of Veterans Affairs and served as an advisor to both agencies on matters related to drug and device information and electronic health records. He is a Fellow of the American College of Physicians and a Fellow of the American College of Medical Informatics.

In February 2013, **May Cheh** retired after more than 32 years of service to NLM and the Lister Hill Center. She earned her MS degree in Computer Science from American University, DC and a BA degree in Chemistry from U. California, Berkeley. She worked as an informatician at the Lister Hill Center and has worked on projects in decision support, including the AI/RHEUM project which she helped to develop and take to clinical evaluation in primary care physicians' offices. She worked on UMLS projects including the Large Scale Vocabulary Test and early development of the UMLS Terminology Server, then, called the Knowledge Source Server. Recently, she headed the Genetics Home Reference Project, which provides a website of genetics information for patients and their families. She has been involved with informatics training at LHC since the 1980's, first as coordinator of the NIH Clinical Elective for Medical Students and then as Director of the LHC's Informatics Training Program since its inception in 1991.

In April 2013, **Aaron Navarro**, **PhD** retired after working in the IT field for nearly forty years. He was a principal contributor on several major projects for DoD, NASA, FAA, USPTO, and NLM including the Space Station program, Air Traffic Control System, US Patent Database, and Informatics R&D. Dr. Navarro joined LHC in 2003 as the Assistant Director for Program Development. He contributed to several Center R&D initiatives and oversaw the software support contracts totaling over 75 contractors. Dr. Navarro holds a PhD from the University of Maryland, an MS from Syracuse University, and a BS, magna cum laude, from the New York Institute of Technology. He is a collector of Chinese snuff bottles and Japanese netsukes. In retirement, if he's not swinging a tennis racket or golf club, you may find him at the bridge table.

In April 2013, **Paul Theerman PhD** retired from Federal service after nearly 15 years of exemplary service at the National Library of Medicine, Head of the Images and Archives Section of the History of Medicine Division,. He will be taking a new position as Associate Director of the new Center for the History of Medicine and Public Health at the New York Academy of Medicine. Dr. Theerman served as head of the Images and Archives Section since December 1998. In that capacity he has recruited all the staff of the Section. Under Paul's leadership, the staff of I&A have worked to expand the digitization initiatives of the Division, including solidifying the Profiles in Science digital manuscripts project, now at over 35 sites; putting manuscript finding aids on line, and expanding their reach through sponsoring a finding aids consortium; and migrating Images from the History of Medicine to a new software platform. Dr. Theerman was instrumental in acquiring many manuscript and image collections. He also was part of the teams that organized a number of historical conferences and exhibits.

Awards

The NLM Board of Regents Award for Scholarship or Technical Achievement is awarded to recognize and stimulate independent creativity leading to scholarly and/or technical achievements that enrich biomedicine. The recipients of the 2013 award were **Dr. Sameer K. Antani and Dr. Dina Demner-Fushman** for leading the underlying research, design, development, and public launch of the OpenI system.

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 2013 award were **Dr. Donna R. Maglott** in recognition of her accomplishments in creating ClinVar, a database that will benefit both researchers and patients by establishing a single resource for documenting human genetic variations and their clinical relevance, and **Dr. Rebecca J. Williams** in recognition of her extraordinary service to the ClinicalTrials.gov program.

The NLM Director's Honor Awards, presented in recognition of exceptional contributions to the NLM mission, were awarded to **Diane L. Boehr** in recognition of substantial and innovative contributions to the NLM cataloging operation and its impact on libraries worldwide, and **Dr. Sally E. Howe** for her contributions in the complex arena of interagency program development that have advanced the use of high performance computing and communications for the benefit of biomedical research and the public's health.

The Phillip C. Coleman Award was presented to **Susan C. Moffatt**, for her demonstrated outstanding leadership, management, and mentoring skill in assisting employees to achieve their highest career potential.

The NIH Merit Award was presented to six individuals including **Dr. Kathel Dunn** for exemplary mentoring and workforce development contributions that enrich the individuals, the NLM mission, and the library and information services profession, **Dr. Stephen J. Greenberg** for outstanding service to the research community through his commitment to and knowledge of medical librarianship and medical history, **Jennifer R. Heiland-Luedtke** for exceptional support of NLM's customers by measuring and improving user experiences of NLM websites, **Mary W. Herron** for outstanding public service to NLM's users through her leadership and management of the *NLM Technical Bulletin*, **David L. Nash** for sustained outreach efforts to minority populations that have raised awareness and use of NLM information services and promoted interest in scientific and health careers, and **Dr. Barbara A. Rapp** for superior service and leadership in program planning and evaluation at the National Library of Medicine.

The NIH Director's Award was presented to one NLM staff member as an individual award and five NLM staff as part of NIH group awards. **Dr. Terry S. Yoo** received an individual award for development and maintenance of the Insight Toolkit (ITK) as the open source standard for 3-dimensional image segmentation and registration, and associated benefits to patients. The Office of the Director, NIH Genomic Data Sharing Oversight and Governance Committees Group Award included **Michael L. Feolo, Dr. James M. Ostell, Dr. Stephen T. Sherry, and Lora Ziyabari** in recognition of exceptional dedication in coordinating and overseeing the NIH Policy for Sharing Data Obtained in NIH Supported or Conducted Genome-wide Association Studies. The Office of Director, Enhancing Peer Review Survey Consultant Group Award included **Dr. Zoe E. Huang** in recognition of the outstanding collaborations of the Enhancing Peer Review Consultant Group to implement the second phase of Continuous Review of Peer Review.

NLM Staff were recognized for their outstanding efforts by the White House and HHS as follows.

The White House Champions of Change "Open Science" Award was presented to **Dr. David J. Lipman** for his outstanding work in promoting and using open scientific data and publications to accelerate progress and improve our world.

The HHS*ignite* Awards were presented to two teams including **Hua Florence Chang, Ziqiang Dong, David S. Hale, Quynh-Thu D. Nguyen, and Ying Sun** for the *Pillbox Takes the Red Pill: Promoting innovation and developer engagement around HHS medication data through open source.* The objective is to build an active community of developers through open source to create innovative products, apps, and services around drug identification, reference, and image data sets. The HHS*ignite* Award was also presented to **Dr. Terry S. Yoo and Dr. David T. Chen** (Contractor) for *The NIH 3D Printing Exchange.* The objective is to create an online portal where researchers and educators can download files of biomedical structures that can be printed on a desktop 3D printer or via an online or in-store 3D print service provider.

NLM Diversity Council

The Director's Employee Education Fund, coordinated by the NLM Diversity Council, has proven to be a successful continuing education mechanism for NLM employees. The fund was established in 1998 by the NLM Director, Dr. Donald A.B. Lindberg to empower every employee with the ability to reach his/her fullest potential. In FY 2013, the fund approved 29 staff to take 48 courses, the majority of which were undergraduates. The school with the largest number of NLM enrollees was the University of Maryland (20), followed by Montgomery College (4). Other institutions included: Catholic University, American University, FAES-NIH Campus, College of Notre Dame of Maryland, Walden University, USA Graduate School, Doshisha University, Syracuse University, Kent State University, Strayer University, San Jose State University and Oregon State University.

Course disciplines included: Computer Security, Software/Hardware Assurance and Security Management, Project Management, Marketing Management and Innovation, Research Methods for Libraries and Information Centers, Aspects of Financial Reporting, Financial Analysis, Sociology, Intercultural Communities and Leadership, Human Resource Management, Public Health Program Planning and Evaluation, Social and Behavioral Sciences, History and Theory of Cultural Heritage Institutions, History and Ethics of Biotechnology, among others. Students enrolled in traditional classroom instructions as well as online courses. The Diversity Council continues its efforts to publicize the availability of the fund.

NLM 2013 Health Expo/Take Your Child to Work Day

On April 25, 2013, the NLM Diversity Council sponsored its "Healthy Lifestyles for You and Your Family Expo" in conjunction with NIH's "Take Your Child to Work Day".

The NLM Expo and events featured a full day's worth of educational, family-oriented activities focusing on healthy living styles for both adults and children, with a multi-cultural theme. The Emmy Award-winning Food-Play Productions which uses both live theatre and interactive media to promote healthy eating and exercise activities were joined this year by the NLM sponsored events: "Treat Your Body Right" Health and Fitness Show, which illustrated sensible and well-balanced food choices. The Health Expo also featured Mad Scientist Taste and Smell Workshop, Mad Scientist Cells Workshop, Distance Learning Demonstrations, a tour of the Library, the Visitors Center and a Scavenger Hunt of NLM's current exhibition, *Native Voices: Native Peoples' Concepts of Health and Illness.* The program concluded with the audience participating in "YOJOROBICS!"

Appendix 4: Regional Medical Libraries

1. **MIDDLE ATLANTIC REGION**
Middle Atlantic Region
University of Pittsburgh
Health Sciences Library System
200 Scaife Hall, 3550 Terrace Street
Pittsburgh, Pennsylvania 15261
Phone: (412) 648-2065 Fax: (412) 624-1515
States served: DE, NJ, NY, PA
URL: http://nnlm.gov/mar

2. **SOUTHEASTERN/ATLANTIC REGION**
University of Maryland at Baltimore
Health Science and Human Services Library
601 Lombard Street
Baltimore, MD 21201-1583
Phone: (410) 706-2855 Fax (410) 706-0099
States served: AL, FL, GA, MD, MS, NC, SC, TN, VA, WV, DC, VI, PR
URL: http://nnlm.gov/sea/

3. **GREATER MIDWEST REGION**
University of Illinois at Chicago
Library of the Health Sciences (M/C 763)
1750 West Polk Street
Chicago, IL 60612-4330
Phone: (312) 996-2464 Fax (312) 996-2226
States served: IA, IL, IN, KY, MI, MN, ND, OH, SD, WI
URL: http://nnlm.gov/gmr

4. **MIDCONTINENTAL REGION**
University of Utah
Spencer S. Eccles Health Sciences Library
10 North 1900 East
Salt Lake City, Utah 84112-5890
Phone: (801) 587-3412 Fax: (801) 581-3632
States served: CO, KS, MO, NE, UT, WY
URL: http://nnlm.gov/mcr

5. **SOUTH CENTRAL REGION**
Houston Academy of Medicine
Texas Medical Center Library
1133 MD Anderson Boulevard
Houston, TX 77030-2809
Phone: (713) 799-7880 Fax: (713) 790-7030
States served: AR, LA, NM, OK, TX
URL: http://nnlm.gov/scr

6. **PACIFIC NORTHWEST REGION**
University of Washington
Health Sciences Libraries and Information Center
Box 357155
Seattle, WA 98195-7155
Phone: (206) 543-8262 Fax: (206) 543-2469
States served: AK, ID, MT, OR, WA
URL: http://nnlm.gov/pnr

7. **PACIFIC SOUTHWEST REGION**
University of California, Los Angeles
Louise M. Darling Biomedical Library
Box 951798
Los Angeles, CA 90025-1798
Phone: (310) 825-1200 Fax: (310) 825-5389
States served: AZ, CA, HI, NV and US Territories in the Pacific Basin
URL: http://nnlm.gov/psr

8. **NEW ENGLAND REGION**
University of Massachusetts Medical School
Lamar Soutter Library
55 Lake Avenue, North
Worcester, MA 01655 – 0121
Phone: 508-856-5979
Fax: 508-856-5977
States Served: CT, MA, ME, NH, RI, VT
URL:http://nnlm.gov/ner

Appendix 5: Board of Regents

The NLM Board of Regents meets three times a year to consider Library issues and make recommendations to the Secretary of Health and Human Services affecting the Library.

Chairperson

Evens, Ronald G., MD
Mallinckrodt Professor of Radiology
Washington University of Medicine
Saint Louis, MO 63110

Members

Fleming, David A., MD, MA, FACP
Professor and Chairman
Department of Internal Medicine
Director, MU Center for Health Ethics
University of Missouri School of Medicine
Columbia, MO 65212

Gottlieb, Katherine, MBA, DPS (h.c)
President and CEO
Southcentral Foundation
Anchorage, AK 99508

Greenes, Robert A., MD, PhD
Professor and Ira A. Fulton Chair in Biomedical Informatics
College of Health Solutions
Arizona State University
Phoenix, AZ 85004

Lewis, Henry, PharmD
Tallahassee, Florida 32303

MacKay, Trudy, PhD
Distinguished University Professor of Genetics
Department of Genetics
North Carolina State University
Raleigh, NC 87695-7614

Mitchell, Joyce A., PhD
Professor and Chair and Associate Vice President
Department of Biomedical Informatics
University of Utah School of Medicine
Salt Lake City, UT 84112

Roskies, Ralph Z., PhD
Professor of Physics, University of Pittsburgh
Scientific Director, Pittsburgh Supercomputing Center
Pittsburgh, PA 15213

Ryan, Mary L., MLS, MPH
UMSAS Library Director/Professor
University of Arkansas for Medical Sciences Library
Little Rock, AK 72205-7186

Scutchfield, F. Douglas, MD
Peter P. Bosomworth Professor of Health Services Research and Policy
University of Kentucky College of Public Health
Lexington, KY 40536-0003

Yokote, Gail A., MS
Associate University Librarian for Sciences and Technical Services
Peter J. Shields Library
University of California, Davis
Davis, CA 95616

Appendix 6: Board of Scientific Counselors, Lister Hill Center for Biomedical Communications

The Board of Scientific Counselors (BSC) provides advice on NLM's intramural research and development programs for the Lister Hill Center for Biomedical Communications.

Chairperson

Mandl, Kenneth D., MD
Professor
Harvard Medical School
Director, Intelligent Health Laboratory
Children's Hospital Informatics Program
Children's Hospital Boston
Boston, MA 02115

Members

Berner, Eta S., EdD
Professor, Health Informatics for Patient Safety/Quality
Department of Health Services Administration
School of Health Professions
Professor, Department of Medical Education
School of Medicine
University of Alabama at Birmingham
Birmingham, AL 35294

Chapman, Wendy W., PhD
Chair
Department of Biomedical Informatics
University of Utah
Salt Lake City, UT 84112-5750

Hammond, William E., PhD
Director
Duke Center for Health Informatics
Duke University
Durham, NC 27705

Hicks, LeRoi S., MD
Associate Professor
Division of Hospital Medicine
UMass Memorial Health Care
Worcester, MA 01605

Horii, Steven C., MD
Professor of Radiology and Clinical Director
Medical Informatics Group
Department of Radiology
Hospital of the University of Pennsylvania
Philadelphia, PA 19104

Rucker, Donald W., MD
Chief Operating Officer, IDEA Studio
Associate Dean for Innovation
Professor of Clinical Emergency Medicine and Biomedical Informatics
Ohio State University
Columbus, OH 43210

Were, Martin C., MD
Assistant Professor of Medicine
Department of General Internal Medicine and Geriatrics
Indiana University School of Medicine
Indianapolis, IN 46202-3012

Appendix 7: Board of Scientific Counselors, National Center for Biotechnology Information

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

Chairperson

Seidman, Christine E., MD
T.W. Smith Professor of Medicine and Genetics
Harvard Medical School
Boston, MA 02115

Members

Babbitt, Patricia C., PhD
Professor
Department of Bioengineering & Therapeutic Sciences
University of California, San Francisco
San Francisco, CA 94158-2330

Benham, Craig J., PhD
Professor
Department of Biomedical Engineering
UC Davis Genome Center
University of California, Davis
Davis, CA 95616

De Crecy-Lagard, Valerie A., PhD
Associate Professor
Department of Microbiology
University of Florida
Gainesville, FL 32611

Edwards, Scott V., PhD
Professor
Department of Organismic and Evolutionary Biology
Harvard University
Cambridge, MA 02138

Guy, R. Kiplin, PhD
Assistant Professor
Department of Chemical Biology and Therapeutics
St. Jude Children's Research Hospital
Memphis, TN 38105

Lee, Christopher J., PhD
Professor
Department of Chemistry & Biochemistry
Molecular Biology Institute
University of California, Los Angeles
Los Angeles, CA 90095-1570

Wu, Chung-I, PhD
Professor
Department of Ecology and Evolution
University of Chicago
Chicago, IL 60637

Appendix 8: Biomedical Library and Informatics Review Committee

The Biomedical Library and Informatics Review Committee meets three times a year to review applications for grants under the Medical Library Assistance Act.

Chairperson

Subramaniam, Shankar, PhD
Professor and Chair
Department of Engineering
University of California, San Diego
La Jolla, CA 92093-0412

Members

Alpi, Kristine M., MLS, MPH
Director & Adjunct Assistant Professor
Department of Population Heath & Pathobiology
Veterinary Medicine Library
North Carolina State University
Raleigh, NC 27607

Bennett, Kristin P., PhD
Professor
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY 12180

Butte, Atul J., MD, PhD
Associate Professor
Stanford University
Systems Medicine/Pediatrics
Stanford, CA 94305-5415

Cardozo, Timothy J., MD, PhD
Associate Professor
Department of Biochemistry and Molecular Pharmacology
New York University School of Medicine
New York, NY 10016

Craven, Mark W., PhD
Professor
Department of Biostatistics and Medical Informatics
University of Wisconsin-Madison
Madison, WI 53706

Gorman, Paul N., MD
Associate Professor
Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR 97239

Hurdle, John F., MD, PhD
Associate Professor
Department of Biomedical Informatics
University of Utah School of Medicine
Salt Lake City, UT 84112

Kalet, Ira J., PhD
Professor Emeritus
Radiation Oncology Department
Department of Biomedical Informatics and Medical Education
University of Washington
Seattle, WA 98195-6043

Kho, Abel N., MD
Assistant Professor
Division of General Internal & Biomedical Informatics
Northwestern University
Chicago, IL 60611

Kochi, Julia K., MILS
Director, Digital Library & Collections
Library & Center for Knowledge Management
University of California, San Francisco
San Francisco, CA 94143-0840

Koonce, Taneya Y., MPH, MSLS
Associate Director of Research
Knowledge Management/Eskind Biomedical Library
Vanderbilt University Medical Center
Nashville, TN 37323-8340

Larson, Ray R., PhD
Professor
School of Information
University of California, Berkeley
Berkeley, CA 94720

Lussier, Yves A., MD
Professor of Medicine
Associate Vice President for Health Sciences
University of Arizona
Tucson, AZ 85720

Payne, Philip, PhD
Associate Director for Data Sciences
Chair, Department of Biomedical Informatics

Center for Clinical and Translational Science
The Ohio State University
Columbus, OH 43210

Pestian, John P., PhD
Professor, Pediatrics
Director, Computational Medicine Center
Cincinnati Children's Hospital Medical Center
University of Cincinnati
Cincinnati, OH 45229

Rios, Gabriel R., MLS
Deputy Director
Lister Hill Library
University of Alabama at Birmingham
Birmingham, AL 35294-0013

Savova, Guergana K., PhD
Assistant Professor
Children's Hospital Boston
Harvard Medical School
Boston, MA 02115

Sonnenberg, Frank A., MD
Professor of Medicine
Rutgers Robert Wood Johnson Medical School
University of Medicine and Dentistry of New Jersey
Brunswick, NJ 08903-0019

Szolovits, Peter, PhD
Professor
Computer Science and Artificial Intelligence Lab
Massachusetts Institute of Technology
Cambridge, MA 02139

Veinot, Tiffany, CE, PhD
Assistant Professor
School of Information
University of Michigan
Ann Arbor, MI 48109-1285

Weng, Chunhua, PhD
Assistant Professor
Department of Biomedical Informatics
Columbia University
New York, NY 10032

Appendix 9: Literature Selection Technical Review Committee

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

Chairperson

Ogunyemi, Omolola, PhD
Director, Center for Biomedical Informatics
Associate Professor, Medical Sciences Institute
Charles Drew University of Medicine and Science
Los Angeles, CA 90059

Members

Cabello, Felipe C., MD
Professor
Department of Microbiology & Immunology
New York Medical College
Valhalla, NY 10595

Cheung, Dorothy S., MD
Assistant Professor
MACC Fund Research Center
Milwaukee, WI 53226

Courtney, Karen L., PhD
Assistant Professor
School of Health Information Science
Victoria, BC, Canada, V8W 3P5

Crummett, Courtney, MS, MLS
Bioinformatics and Biosciences Librarian
Massachusetts Institute of Technology Libraries
Cambridge, MA 02139

Delclos, George L., MD, PhD
Professor
Division of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health
University of Texas-Houston
Houston, TX 77030

Gwinn, Marta, MD
Senior Consultant
Office of Public Health Economics
Centers for Disease Control and Prevention
Atlanta, GA 30333

Joe, Jennie R., PhD, MPH, MA
Professor Emeritus
Family and Community Medicine
University of Arizona College of Medicine
Tucson, AZ 85724

Nguyen, Thu Annelise, PhD
Associate Professor
Department of Diagnostic Medicine
Kansas State University
Manhattan, KS 66506

Pascoe, John M., PhD
Professor
Department of Pediatrics
Wright State University
Dayton, OH 45404

Pascoe, John R., BVSc, PhD
Professor and Executive Associate Dean
Dean's Office
School of Veterinary Medicine
University of California, Davis
Davis, CA 95616

Phillips, William R., MD
Theodore J. Phillips endowed Professor in Family Medicine
Department of Family Medicine
University of Washington
Seattle, WA 98195

Rayo, Jaya, MD
Deputy Editor, Annals Internal Medicine
American College of Physicians
Philadelphia, PA 19106

Tannery, Nancy H., MLS
Senior Associate Director
Health Sciences Library System
University of Pittsburgh
Pittsburgh, PA 15261

Appendix 10: PubMed Central National Advisory Committee

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

Chairperson

Thibodeau, Patricia L., MLS, MBA
Associate Dean for Library Services & Archives
Medical Center Library
Duke University
Durham, NC 27710

Members

Anderson, Ivy, MLS
Director
Collection Development & Management Program
California Digital Library
Oakland, CA 94612

Bedard, Martha A., MSLS, MA
Vice Provost for University Libraries
University of Connecticut
Storrs, CT 06269

Bourne, Philip E., PhD
Professor
Department of Pharmacology
Skaggs School of Pharmacy and Pharmaceutical Sciences
University of California, San Diego
La Jolla, CA 92092-0505

Cantu, Adelita G., PhD
Consulting Associate Professor
Department of Psychiatry and Behavioral Sciences
Stanford University Medical School
San Francisco, CA 94123

Colamarino, Sophia A., PhD
Consulting Associate Professor
Department of Psychiatry and Behavioral Sciences
Stanford University Medical School
San Francisco, CA 94123

Courant, Paul N., PhD
University Librarian and Dean of Libraries
University of Michigan Library
University of Michigan
Ann Arbor, MI 48109-1205

Engelward, Bevin P., PhD, Sc.D.
Associate Professor
Department of Biological Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Fassler, Jan S., PhD
Professor
Department of Biology
University of Iowa
Iowa City, IA 52242

Haricombe, Lorraine J., PhD
Dean of Libraries
Watson Library
The University of Kansas
Lawrence, KS 68045

Jongeneel, Cornelis V., PhD
Director
High-Performance Biological Computing Group
National Center for Supercomputing Applications
Department of Bioengineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Meglio, Delores, MS
Digital Publishing Consultant
Sways DM LLC
Brooklyn, NY 11234

Morse, Randall H., PhD
Research Scientist
Molecular Genetics Program
Wadsworth Center
Albany, NY 12201

Rossner, Mike, PhD
Executive Director
The Rockefeller University Press
New York, NY 10021

Terry, Sharon F., MA
President and CEO
Genetic Alliance
Washington, DC 20008-2304

Appendix 11: Organizational Acronyms and Initialisms Used in this Report

Acronym	Meaning of Acronym
AAHSL	Association of Academic Health Sciences Libraries
AABB	Non-profit association formerly known as American Association of Blood Banks
AAPA	American Academy of Physicians Assistants
ABC	Advanced Biomedical Tele-Collaboration Test Bed)
ACIOP	AIDS Community Information Outreach Program
ACLA	American Clinical Laboratory Association
ACORN	Automatically Creating OLDMEDLINE Records for NLM
ACP	American College of Physicians
ACSI	American Customer Satisfaction Index
AFIP	Armed Forces Institute of Pathology
AG	Access Grid
AHIC	American Health Information Community
AHILA	Association for Health Information and Libraries in Africa
AHRQ	Agency for Healthcare Research and Quality
AIDS*info*	Acquired Immune Deficiency Syndrome*info* (database)
AJPP	African Medical Journal Editors Partnership Program
ALTBIB	Alternatives to Animal Testing
AIH	American Indian Health (Web portal)
AME	Automated Metadata Extraction
AMIA	American Medical Informatics Association
AMPA	American Medical Publishers Association
AMPATH	Academic Model Providing Access to Healthcare
AMWA	American Medical Women's Association
APDB	Audiovisual Program Development Branch
API	Applied Programming Interface
APIRE	American Psychiatric Institute for Research and Education
ARRA	American Recovery and Reinvestment Act
ASCCP	American Society for Cervical Pathology and Colposcopy
ASHG	American Society of Human Genetics
ASPE	Assistant Secretary for Planning and Evaluation, HHS Office of the
ASPR	Assistant Secretary for Preparedness and Response, HHS Office of the
ATC	Anatomic Therapeutic Class
BAC	Bacterial Artificial Chromosome
BarSTool	Barcode Submission Tool
BD2K	Big Data to Knowledge initiative (NIH)
BGMUT	Blood Group Antigen Gene Mutation Database
BHEPP	Bethesda Hospitals' Emergency Preparedness Partnership
BISTI	Biomedical Information Science and Technology Initiative
BITA	Biomedical Image Transmission via Advanced Networks
BLAST	Basic Local Alignment Search Tool
BLIRC	Biomedical Library and Informatics Review Committee
BMT	Boundary Marking Tool
BN	Brand Name
BOR	Board of Regents
BSAT	BMT Study Administration Tool
BoSC	Board of Scientific Counselors
BSD	Bibliographic Services Division
BSE	Book Scan Editor
BSN	Bioinformatics Support Network
C3PI	Computational Photography Project for Pill Identification
CADTH	Canadian Agency for Drugs and Technologies in Health
CAM	Complementary and Alternative Medicine
C&A	Certification & Accreditation (audit)
CARe	Candidate Gene Association Resource project
CAS	Collection Access Section
CBB	Computational Biology Branch
CBIR	Content-Based Image Retrieval
CBRN	chemical, biological, radiological and nuclear (incidents)
CCB	Configuration Control Board
CCDS	Consensus CoDing Sequence
CCHD	Critical Congenital Heart Disease
CCHIT	Commission for Healthcare Information Technology
CCR	Central Contractor Registration
CCRIS	Chemical Carcinogenesis Research Information System
CDART	Conserved Domain Architecture Retrieval Tool
CDD	Conserved Domain Database
cDNA	Complementary DNA
CEB	Communications Engineering Branch
CEL	Affymetrix Cell intensity (file)
CER	Comparative Effectiveness Research
CgSB	Cognitive Science Branch
CHEBI	Chemical Entities of Biological Interest
ChEMBL	Computational Chemical Biology Group database
ChemIDplus	Chemical Identification File
CHEMM	Chemical Hazard Event Medical Management
CHIA	CEB Histology Image Assistant
CHIC	Chickasaw Health Information Center
CHRIS	Consumer Health Resource Information Service

Acronym	Meaning of Acronym
CIN	Cervical Intraepithelial Neoplasia
CIP	Cataloging-in-Publication
CIT	Center for Information Technology
CLML	Current List of Medical Literature
CMAX	Collaborative Multi-Agency eXercise (BHEPP disaster drill)
CMP	Colposcopy Mentorship Program
CMS	Centers for Medicare and Medicaid Services
CMSP	Cooperative Medical Science Program, US-Japan
CMT	Convergent Medical Terminology
COOP	(NIH Pandemic Flu) Continuity of Operations Plan
CORE	Clinical Observations Recording and Encoding
CoreBio	Core Bioinformatics Facility
CounterACT	Countermeasures Against Chemical Threats
CPS	Commercial Peering Service
CPSC	Center for Public Service Communication
CPT	Current Procedural Terminology
CRAC	Computer Room Air Conditioner
CRAH	Computer Room Air Handler
CRD	Centre for Reviews and Dissemination (England)
CRI	Clinical Research Informatics
CRISP	Computer Retrieval of Information on Scientific Projects
CSB	Computer Science Branch
CSI	Commission on Systemic Interoperability
CSIRC	Computer Security Incident Response Center (HHS)
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
CWDM	Coarse Wave Division Multiplexing
CXR	Chest X-ray
DAC	Data Access Committees
DAR	Data Access Request
DARE	Database of Reviews of Effects
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
dbVar	Database of Genomic Structural Variation
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
DELTA-BLAST	Domain Enhanced Lookup Time Accelerated BLAST
DHHS	Department of Health and Human Services
DHS	Department of Homeland Security
DICOM	Digital Imaging and Communications in Medicine
DIMRC	Disaster Information Management Research Center
DIRLINE	Directory of Information Resources Online
DLXS	Digital Library Extension Service
DMSZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen (German Collection of Microorganisms and Cell Cultures)
DNA	Deoxyribonucleic Acid
DOT	Department of Transportation
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
EBM	Evidence-Based Medicine
EBP	Evidence-Based Practice
ECHO	European Community Humanitarian Office
Educollab	Educational Collaborators
EEO	Equal Employment Opportunity
EFTS	Electronic Funds Transfer Service
HER	Electronic Health Record
EMBL	European Molecular Biology Laboratory
EMR	Electronic Medical Record
EMS	Emergency Medical Services
EnHIP	Environmental Health Information Partnership
EnHIOP	Environmental Health Information Outreach Program
EP	Extramural Programs
EPA	Environmental Protection Agency
eQTL	expression quantitative trait loci
eRA	Electronic Research Administration
ERG	Emergency Response Guidebook
ESI	Early Stage Investigators
EST	Expressed Sequence Tag
ETIC	Environmental Teratology Information Center
eTK	Electronic Thorndike and Kibre

Acronym	Meaning of Acronym
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
FMS	Facilities Management Section
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
GIA	Gene Indexing Assistant
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
HAVnet	Haptic Audio Video Network for Education Technology
HBCU	Historically Black Colleges and Universities
HCN	Hispanic Communications Network
HD	High Definition
HGVS	Human Genome Variation Society
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

Acronym	Meaning of Acronym
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
HUMLO	Hemoglobinopathies Uniform Medical Language Ontology Project (NHLBI)
I3	Image Indexing Initiative
IAIMS	Integrated Advanced Information Management Systems
IBIS	Inferred Biomolecular Interactions Server
ICC	Incident Command Center
ICCVAM	Interagency Coordinating Committee on The Validation of Alternative Methods
ICD	International Classification of Diseases
ICMJE	International Committee of Medical Journal Editors
ICs	Institutes and Centers (of NIH)
ICT	Information and Communication Technologies
ICU	Intensive Care Unit
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)
*info*SIDA	infoSíndrome de Inmunodeficiencia Adquirida (database)
IOC	Indicator of Compromise
IP	Interactive Publications
IPv6	Next Generation Internet, Version 6
IRB	Institutional Review Board
IRC	In-Row Coolers

Acronym	Meaning of Acronym
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
ITSC	Information Technology Service Center
ITSMS	Information Technology Service Management System
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
LANDHI	Latin American Disaster and Health Information Network
LC	Library of Congress
LHI	Leading Health Indicators (HHS)
LHNCBC	Lister Hill National Center for Biomedical Communications
LID	Laboratory for Informatics Development
LIPID MAPS	Lipid Metabolites and Pathways Strategy
LITE	Librarian Infobutton Tailoring Environment
LJI	List of Journals Indexed
LO	Library Operations
LOINC	Logical Observations Identifiers, Names, Codes
LPF	Lost Person Finder
LRG	Locus Reference Genomic
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
MEO	Medical Education and Outreach
MEPI	Medical Education Partnership Initiative (Africa)
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
MIR	Mandiant Intelligent Response
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
MTIFL	Medical Text Indexer First Line
MTMS	MeSH Translation Management System
MUSC	Medical University of South Carolina
NAC	Network Access Control
NA-MIC	National Alliance of Medical Image Computing
NAML	Network of African Medical Libraries
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NCATS	National Center for Advancing Translational Sciences
NCBC	National Center for Biomedical Computing
NCBI	National Center for Biotechnology Information
NCCS	NIHI Consolidated Collocation Site
NCHS	National Center for Health Statistics
NCI	National Cancer Institute
NCRR	National Center for Research Resources
NCVHS	National Committee on Vital and Health Statistics
NDC	National Data Codes
NDF-RT	National Drug File – Reference Terminology
N*e*HC	National *e*-Health Collaborative
NEI	National Eye Institute

Acronym	Meaning of Acronym
NewSTEPs	Newborn Screening Technical assistance and Evaluation Program
NGI	Next Generation Internet
NHANES	National Health and Nutrition Examination Surveys
NHGRI	National Human Genome Research Institute
NHIN	National Health Information Network
NHLBI	National Heart. Lung, and Blood Institute
NIA	National Institute on Aging
NIAID	National Institute of Allergy and Infectious Diseases
NIBIB	National Institute of Biomedical Imaging and Bioengineering
NICHD	National Institute of Child Health and Human Development
NICHSR	National Information Center on Health Services Research and Health Care Technology
NIDCD	National Institute on Deafness and other Communication Disorders
NIDCR	National Institute of Dental and Cranio-facial Research
NIDDK	National Institute of Diabetes, Digestive, and Kidney Diseases
NIEHS	National Institute of Environmental Health Sciences
NIGMS	National Institute of General Medical Sciences
NIH	National Institutes of Health
NIHMS	NIH Manuscript Submission
NIH PI	NIH Pathways to Independence Award
NIMH	National Institute of Mental Health
NIMHD	National Institute on Minority Health and Health Disparities
NINDS	National Institute of Neurological Disorders and Stoke
NIOSH	National Institute for Occupational Safety and Health
NIS	Network Information Service
NIST	National Institute of Standards and Technology
NLM	National Library of Medicine
NLM LitArch	NLM Literature Archive
NLP	National Language Processing System
NMAETC	National Minority AIDS Education and Training Center
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)
NQF	National Quality Forum
NRCBL	National Reference Center for Bioethics Literature
NSF	National Science Foundation
NTCC	National Online Training Center and Clearinghouse
OA	Open Access
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
ODS	Office of Dietary Supplements (NIH)
OERC	Outreach Evaluation Resource Center
OHA	Office of Health Affairs (DHS)
ORF	Original Release Format
OHIPD	Office of Health Information Programs Development
OMB	Office of Management and Budget
OMIA	Online Inheritance in Animals (database)
OMIM	Online Mendelian Inheritance in Man (database)
OMSSA	Open Mass Spectrometry Search Algorithm
ONC	Office of National Coordinator (for Health Information Technology)
OPASI	Office of Portfolio Analysis and Strategic Initiatives
OPD	OLDMEDLINE Serials Application
OPD	Outreach Products Database
ORDR	Office of Rare Diseases and Research (NIH)
OSA	Optical Society of America
OTB	Orfeo Toolbox
ORWH	Office of Research on Women's Health
OSIRIS	Open Source Independent Review and Interpretation System
PA	Pharmacological action
PAHO	Pan American Health Organization
PAL	Potential Abstract Labels (tool)
PBM	Pharmacy Benefit Manager
PCA	Personal Computer Advisory Committee
PCOR	Patient Centered Outcomes Research
PCR	Polymerase Chain Reaction
PDA	Personal Digital Assistant
PDRS	Publisher Data Review System
PDB	Protein Data Bank
PDF	Portable Document Format
PDL	Personal Digital Library
PDM	Patient Data Management
PDQ	Physician Data Query
PFIF	Person Finder Interchange Format

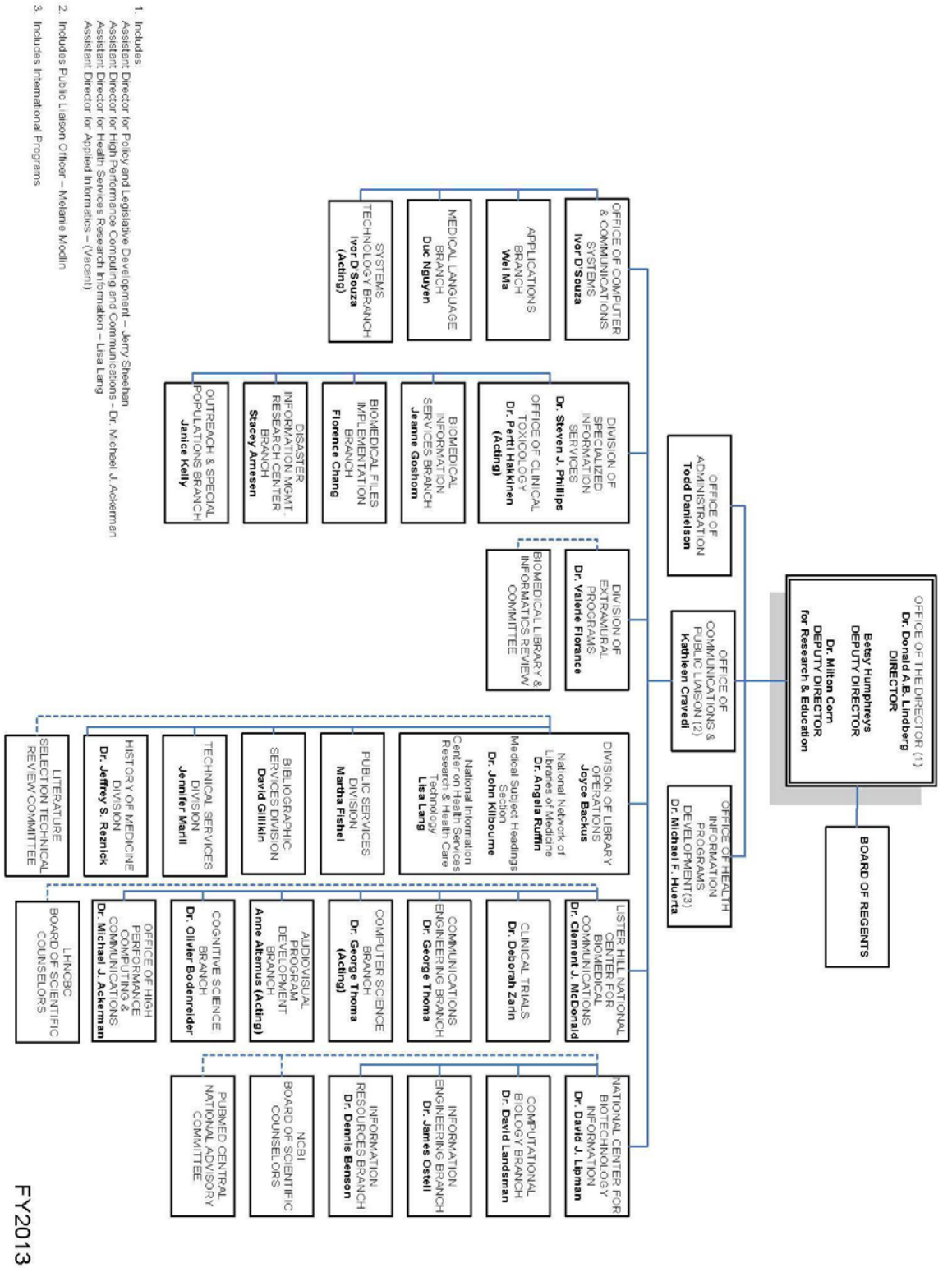
Acronym	Meaning of Acronym
PheGenI	Phenotype-Genotype Integrator
PHLIP	Public Health Law Information Project
PHII	Public Health Informatics Institute
PHP	Public Health Partners
PHR	Personal Health Record
PHS	Public Health Service
PI	Pathway to Independence award
PI	Principal Investigator
PII	Personally Identifiable Information
PICO	Patient/Population, Intervention, Comparison, and Outcome
PID	Pathway Interaction Database (NCI)
PIN	Precise Ingredient (term type), RxNorm
PIV	Personal Identify Verification
PL	Person Locator
PLAWAR*e*	Programmable Layered Architecture With Artistic Rendering
PMC	PubMed Central
PMCI	PubMed Central International
PMC ID	PubMed Central Identification (number)
PNG	Portable Network Graphics (file format)
PRS	Protocol Registration System
PSD	Public Services Division
PubMedHh	PubMed for Handhelds
PUE	Power Use Efficiency
PUG	PubChem Power User Gateway
QA	Quality Assurance
QCIM	Quarterly Cumulative Index Medicus
RAC	Real Application Clusters
RACCE	Residents' Assessment of Competency in Colposcopy Exam
RCDC	Research Condition and Disease Categorization
RCSB	Research Collaboratory for Structural Bioinformatics
RDMS	Rare Disease Maintenance System
RefSeq	Reference Sequence (database)
RefSNP	Reference SNP (database)
RELACIGER	Red Latinoamericana de Centros de Información en Gestión del Riesgo de Desastres (Latin American Network of Risk Management Centers)
REMM	Radiation Even Medical Management
RF2	Release Format version 2
RFA	Request for Applications
RFID	Radio Frequency Identification
RFP	Request for Proposals
RHIN	Refugee Health Information Network
RIDeM	Repository for Informed Decision Making
RML	Regional Medical Library
RNA	Ribonucleic Acid
RNAi	RNA Interference
RPS-BLAST	Reversed Position Specific BLAST
RQS	Request Submission and Tracking System
RRF	Rich Release Format
RSS	Really Simple Syndication
RTECS	Registry of Toxic Effects of Chemical Substances
RTLS	Real Time Location System
RVDS	Remote Virtual Dialogue System
RWD	Responsive Web Design
RWJF	Robert Wood Johnson Foundation
SAB	Source Abbreviations
SBIR	Small Business Innovation Research
SCID	Severe Combined Immunodeficiency
SCR	(MeSH) Supplemental Chemical Records
SDK	Software Development Kit
SEF	Serials Extract File
SEIM	Security Event and Information Management System
SEO	Search Engine Optimization
SEP	Special
SEPs	Special Emphasis Panels
SEQ	Structured Evidence Queries
SIDA	Swedish International Development Agency
SIG	Special Interest Group
SII	Scalable Information Infrastructure
SIS	Specialized Information Services
SKR	Semantic Knowledge Representation
SMART	Scalable Medical Alert and Response Technology
SNOMEDCT	Systematized Nomenclature of Medicine Clinical Terms
SO	Signing Official
SOAP	Simple Object Oriented Protocol (formerly Simple Object Access Protocol)
SOC	(HHS) Secretary's Operation Center
SPER	System for the Preservation of Electronic Resources
SPIN	Shared Pathology Informatics Network
SPIRS	Spine Pathology Image Retrieval System
SPL	Structured Product Labels (FDA)
SPWG	Special Projects Workgroup (NIH)
SRA	Short Read Archive
SRS	Substance Registration System (FDA)
STB	Systems Technology Branch
STEM	Science, Technology, Engineering and Math
STR	Short Tandem Repeat
STTP	Short-Term Trainee Program
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
TICAP	Trusted Internet Connection Access Partners

Acronym	Meaning of Acronym
TIE	Telemedicine Information Exchange
TIFF	Tagged Image File Format
TKMT	Traditional Korean Medical Terms
TILE	Text to Image Linking Engine
TIOP	Toxicology Information Outreach Project
TOXLINE	Toxicology Information Online
TOXNET	Toxicology Data Network
TPA	Third Party Annotation (database)
TMS	Track Management System (NCBI)
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
UCUM	Unified Codes for Units of Measure
UID	Unique Identifier (PubMed)
UIMA	Unstructured Information Management Architecture
UKPMC	United Kingdom PubMed Central
UMLS	Unified Medical Language System
UMLSKS	UMLS Knowledge Source Server

Further information about the programs described in this Administrative report is available from:

Office of Communications and Public Liaison
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894
301-496-6308
E-mail: publicinfo@nlm.nih.gov
Web: www.nlm.nih.gov

Cover:

This year's cover captures the sweep of medical history, from Andreas Vesalius (1514–1564), best known for changing the field of medical research with his groundbreaking book, *De Humani Corporis Fabrica Libri Septem (Seven Chapters on the Structure of the Human Body)*, published in 1543, through the NLM Visible Human Project of the 1990s and into the era of genomics. The US National Library of Medicine collects, organizes and disseminates it all, and even has holdings back to the 11th century BCE.

NLM Organization Chart