

NATIONAL INSTITUTES OF HEALTH

National Library of Medicine

Programs and Services

Fiscal Year 2014

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 Digital Biomedicine	4
Library Operations	6
Program Planning and Management.....	6
Collection Development and Management.....	6
Preservation and Collection Management	8
Vocabulary Development and Standards.....	9
Bibliographic Control.....	11
Information Products	12
Direct User Services	13
Outreach and Training	14
Specialized Information Services	23
Toxicology and Environmental Health Resources	23
Mobile Resources and Apps	26
Disaster Information.....	26
Outreach.....	28
Environmental Health Information Partnership (EnHIP)	28
HIV/AIDS	32
Lister Hill National Center for Biomedical Communications	35
Biomedical Imaging, Multimedia, and 3D Imaging.....	35
Natural Language Processing and Text Mining.....	44
Clinical Vocabulary Standards and Associated Tools.....	48
Next-Generation Electronic Health Records to Facilitate Patient-Centric Care, Clinical Research, and Public Health	50
Information Resource Delivery for Researchers, Care Providers, and the Public.....	52
Disaster Information Management: Lost Person Finder.....	54
Video Production, Retrieval, and Reuse Project.....	55
National Center for Biotechnology Information.....	59
Highlights of the Year	59
Mission	59
Molecular Biology Information Resources.....	60
dbGaP: Integration of Clinical and Genetic Data	68
Entrez Retrieval System	69
Literature Information Resources	69
Research.....	71
Bioinformatics Training and Support.....	71
Extramural Programs.....	73
Overview of FY2014	73
Highlights	75
Research Grants	77
Resource Grants	78
Training and Career Awards	78
Pan-NIH Projects and Interagency Collaborations	80
EP Web Site	81

Office of Computer and Communications Systems	88
Controlled Medical Vocabularies	88
Consumer and Public Health	90
IT Infrastructure Services	91
IT Web Platform Services.....	96
Medical Literature Support & Document Delivery Services.....	96
Outreach and Customer Services	97
Administration.....	98
Financial Resources and Allocations, FY2014	98
Personnel.....	98
NLM Diversity Council.....	101
NLM 2014 Health Exposition.....	101

NLM Organization Chart (Inside back cover)

Appendixes

1. RFA/PA Actions in FY2014 - NLM’s Core Active Grant Programs	82
2. RFA/PA Actions in FY2014 - Multi-Institute Active Announcements in which NLM Participates.....	82
3. NLM New FY2014 Awarded Grants	84
4. FY2014 Grant Budget by NIH Mechanism	87
5. FY2014 Grant Budget by NIH Activity Code	87
6. Regional Medical Libraries.....	102
7. Board of Regents.....	103
8. Board of Scientific Counselors/LHNCBC	105
9. Board of Scientific Counselors/NCBI.....	106
10. Biomedical Library and Informatics Review Committee	107
11. Literature Selection Technical Review Committee	109
12. PubMed Central National Advisory Committee	110
13. Organizational Acronyms and Initialisms Used in this Report.....	111

Figures

1. NLM FY2014 Funding by Investment Area	73
2. NLM FY2014 Projects by Investment Area	74
3. FY2014 150 Active NLM Grants by Topic	76
4. FY2014 Projects by Topic Area.....	77
5. Virtualization Program Progress	95

Tables

1. Collections	18
2. Collection Activities.....	19
3. Cataloging and Indexing	20
4. Services to the Public	21
5. Selected Web Resources	22
6. Success Rates—Core NLM Grant Programs, FY2012-2014.....	75
7. Top 10 Journals Where NLM Grantees Published in FY2014	76
8. Financial Resources and Allocations, FY2014	98
9. FY2014 Full-Time Equivalents (Actual)	98

PREFACE

I'm pleased to report that Fiscal Year 2014 was another banner year of accomplishment in many areas. Here's a partial list of the achievements of our talented staff, advisors, and consultants, to enhance access to biomedical and health information for people around the globe:

- In its 25th anniversary year, NCBI received an HHS *Innovates* award for work with the CDC, FDA, and USDA to apply whole genome sequencing (WGS) to investigate outbreaks of foodborne disease.
- MedlinePlus continued to help people find trusted, up-to-date consumer health information. In FY2014, the site had over one billion page views with over one million people visiting the site daily.
- NLM released a new version of TOXNET, improved results submission features in ClinicalTrials.gov, and redesigned DailyMed, *AIDSinfo* (English), and *infoSIDA* (Spanish) for automatic optimal viewing across all devices, from desktop computers to smartphones.
- The Value Set Authority Center (VSAC) was augmented with an Authoring Tool and VSAC User Forum Webinars and tutorials were launched, improving NLM support for developers and users of value sets that define clinical quality measures required for meaningful use of electronic health records.
- The Lister Hill Center collaborated on the development of the NIH 3D Print Exchange, an HHS *Signites* award winner. It was featured at the USA Science and Engineering Festival in Washington, DC and at the White House Maker Faire.
- NLM continued to develop and provide information relevant to disasters of all kinds, this year responding to Typhoon Haiyan in Southeast Asia and to the Ebola outbreak in West Africa.
- The *Native Voices* exhibition extended its reach, opening as a traveling version with Native communities in Alaska, Hawaii, North Dakota, and Oklahoma. A suite of related classroom activities and lesson plans allows knowledge of the exhibition subject matter to be shared with students everywhere.
- Our popular semi-annual biomedical informatics course, long held at the Marine Biological Laboratory in Woods Hole, Massachusetts, has a new home at Georgia Regents University.
- Combining history with current day science, we planted a clone of a clipping from the Tree of Hippocrates on our front lawn, with Greek Ambassador Christos Panagopoulos as special guest, and used DNA from the dying tree it replaced to register the species in the Barcode of Life project.
- With our symposium, *The National Library of Medicine, 1984-2014: Voyaging to the Future*, and its accompanying blog, we reviewed the past 30 years while considering opportunities for the future—all as background for the next NLM long range planning effort.

Thanks to you all for your stellar public service, your creativity, and your dedication.



Donald A.B. Lindberg
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

In preparation for development of the next long range plan, NLM hosted a symposium entitled "The National Library of Medicine 1984-2014: Voyaging to the Future" on May 14, 2014. The purpose of this symposium was to consider the accomplishments and challenges of the past thirty years, particularly in the context of the priorities identified in the planning processes.

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, which remains in effect in parallel with the NIH-wide plan, is available on the NLM Web site.

Outreach and Consumer Health

NLM carries out a diverse set of activities directed at building awareness and use of its products and services by health professionals in general and by particular communities of interest. Considerable emphasis has been placed on reducing health disparities by targeting health professionals who serve rural and inner city areas. NLM also undertakes initiatives specifically devoted to addressing the health information needs of the public.

An NLM-wide Coordinating Committee on Outreach, Consumer Health and Health Disparities (OCHD), chaired by OCHD, plans, develops, and coordinates NLM outreach and consumer health activities. In FY2014, 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.

With NLM support, the Mentoring In Medicine (MIM) Science and Health Career Exploration program reaches seven public and charter schools in New York City with after school instruction 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. Over the past five years, the program has exposed more than 800 minority students to health care career instruction. Program evaluations demonstrate continued impressive gains in health care knowledge. In FY2014, the program was expanded to include three schools in the Washington, DC area, with curriculum modifications to accommodate inclusion of middle school students. 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 FY2014, NLM continued its support for iVIVA! Peer Tutors Program at the South Texas High School for the Health Professions, 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 website for interested faculty, librarians, and students from high schools around the country. MedlinePlus, in English and Spanish, is being emphasized where applicable. The project also includes a health careers tutoring component that involves the students, teachers, and guidance counselors. Peer tutoring has been extended to other magnet high schools in the Lower Rio Grande Valley, where it also serves as an effective health information outreach program. A Web-based practical implementation guide is available to assist other high schools in considering and implementing a peer tutoring program. The information guide and supporting documents 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>.

As an outgrowth of the iVIVA! project, the HOSA (Health Occupations Student Organization, now known as HOSA: Future Health Professionals) proposed a health information partnership with NLM. The suggested focus is a Health Outreach Ambassador Program designed primarily for high school students active in local HOSA chapters around the country. NLM and the NNLM Southeast/Atlantic Regional Medical Library participated in the June 2014 HOSA National Leadership Conference held in Orlando, FL. After review and discussion, NLM and HOSA have embarked on a pilot project to be held in Utah for the upcoming academic year. The NNLM MidContinental Regional Medical Library in Salt Lake City, UT will participate.

OHIPD also supported an after school outreach program through the University of Maryland called HackHealth (<http://hackhealth.umd.edu/>), which strives to increase middle school students' interest in science, technology, engineering, and mathematics (STEM) fields; improve their health and digital literacy skills; and increase their health-related self-efficacy. In FY2014, the HackHealth program was pilot-tested in three Title I school libraries in Prince George's County, Maryland. Participating students select a personally-relevant health topic, learn how to conduct scientific inquiry into health maintenance and/or disease prevention, and function as health information intermediaries for family and friends. Eight learning modules for HackHealth, along with evaluation instruments, were created based on the pilot testing phase and are freely available at

<http://hackhealth.umd.edu/about-us/project-phases/>, so any school librarian can implement the program throughout the country.

Native American Outreach

In FY2014 OHIPD 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 in the Mid-Atlantic area. An estimated 10,000 persons visited the NLM booth over the course of these pow-wows. These activities are yet another way to increase awareness of NLM's health information resources among segments of the Native American community and the general public.

OHIPD also provided support for three projects related to tribal colleges. A project at Cankdeska Cikana Community College, Spirit Lake Nation, Ft. Totten, ND, provided continuing support for improvements at the tribal college library and development of health information-related educational and community outreach activities coordinated by the library. Support for the 24th annual Tribal College Librarians Institute at Montana State University, Bozeman, MT, a week-long professional development program for fifty tribal college librarians, provided training about health issues affecting Native Communities, research efforts to address those issues, consumer health and biomedical information resources, and the approaches for health information outreach and education in their tribal communities. The Healthy Diet/Ethobotany Project at the Sitting Bull Tribal College, Standing Rock Sioux Tribe, FT Yates, ND aims to build capacity on the reservation and at the college to experiment, demonstrate and grow healthy traditional foods and medicinal plants through efforts to save and preserve local varieties of seeds. The population of Standing Rock suffers from some of the highest rates of diabetes and other diet-related health conditions in the nation. Some of these challenges can be alleviated by a return to traditional sources of food and healthy, locally grown foods.

Native Voices Exhibition

OHIPD staff contributed to the outreach, follow-up and enhancement of the NLM exhibition, *Native Voices: Native Peoples' Concepts of Health and Illness*. OHIPD also assisted with the *Native Voices* traveling exhibition pilot test planning and implementation. Pilot tests were successfully conducted at the Cankdeska Cikana Community College, Spirit Lake Dakota Nation, Ft. Totten, ND, October 2013 opening; National Congress of American Indians Mid-Year Conference, Alaska Native Heritage Center, and Southcentral Foundation, Anchorage, AK, June 2014 opening; Queens Medical Center, Honolulu, HI, July 2014 opening; and Chickasaw Nation,

ARTesian Art Gallery & Studios, Sulphur, OK, August 2014.

Web Evaluation

The Internet and World Wide Web play a dominant role in dissemination of NLM information services, and the Web environment in which NLM operates is rapidly changing and intensely competitive. These two factors continue to support the need for a comprehensive and dynamic NLM Web planning and evaluation process. The Web evaluation priorities of the OCHD include both quantitative and qualitative metrics of Web usage, measures of customer perception, and use of NLM Web sites. During FY2014, 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 of a 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 FY2014, 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 FY2014 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 programs strengthen and expand global access to health literature and information. They are intended to strengthen research and healthcare systems in low- and middle-income countries through outreach, education, capacity-building, and information infrastructure development. Current initiatives include: the African Journal Partnership Project (AJPP), the Network of African Medical Librarians (NAML), and the Tanzanian Health Information Specialist Training Program. NLM also contributes expertise to the Fogarty International Center's Medical Education Partnership Initiative in Africa (MEPI), 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.

African Journal Partnership Project

(<http://ajpp-online.org/>)

The AJPP was developed in collaboration with Fogarty International Center in 2003. Its purpose is to build peer-reviewed, medical journal capacity in sub-Saharan Africa by partnering African medical journals with medical journals in the US and UK. This capacity-building program makes important research being carried out in endemic countries available to the world. The journals provide a venue for publishing African research results that address broad public health issues and are particularly relevant to local clinicians, nurses, health-care workers, policy-makers, and general public.

Eight African journals and six Northern journals are participating in the partnership. The partnerships are as follows: *African Health Sciences (AHS)*—*British Medical Journal*; *Ethiopian Journal of Health Sciences (EJHS)*—*Annals of Internal Medicine*; *Malawi Medical Journal (MMJ)*—*JAMA*; *Medical Journal of Zambia (MJZ)*—*New England Journal of Medicine*; *Ghana Medical Journal (GMJ)*—*Sierra Leone Journal of Biomedical Research (SLJBR)*—*The Lancet*; and *Mali Medical (MM)*—*Annales Africaines de Medecine (AAM)*—*Environmental Health Perspectives*. One immediate goal is to strengthen African Journals for acceptance into MEDLINE. At this time, five of the journals are now indexed by NLM.

Network of African Medical Librarians

(<http://karibouconnections.net/medlibafrica/>)

NLM continues its commitment to utilizing and expanding the leadership of a growing Network of African Medical Librarians (NAML) who received training as NLM Associate Fellows (<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. Support for the activities of the broader community of African librarians is also provided via the Association for Health Information and Libraries in Africa (AHILA), which held its biannual meeting in October, 2014 in Tanzania.

OHIPD supports health information resource training provided by NAML librarians at meetings of AHILA and other locations throughout the year. They were recently awarded an "Innovative Libraries and Women in Science" grant to promote Consumer Health Literacy in African Public Libraries from the Elsevier Foundation ([http://www.elsevier.com/connect/elsevier-foundation-awards-\\$600k-to-innovative-libraries-and-women-in-science](http://www.elsevier.com/connect/elsevier-foundation-awards-$600k-to-innovative-libraries-and-women-in-science)). A representative of OHIPD serves on the advisory board for this project.

NAML librarians continue to train faculty and students within their home institutions, and engage in outreach to areas outside of their capital cities through workshops, newsletters, and lunchtime training sessions for staff. Several have developed institutional repositories with NLM/OHIPD support that can be accessed online from anywhere. They have also 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.” The manual, which is available free of charge, is comprised of seven modules: Information Sources, Searching Tools, Electronic Information Searching Techniques, Intellectual Property Rights, Management of Information, Evaluating Electronic Resources, and Scholarly Communications: (http://karibouconnections.net/medlibafrica/training_module/index.html)

Tanzanian Health Information Specialist Training Program

OHIPD is supporting the development of a curriculum to train medical information specialists for the Tanzanian Ministry of Health and Social Welfare (MOHSW). In January, 2014, MOHSW, OHIPD, and an expert faculty member from the University of Pittsburgh iSchool organized a fact-finding mission to solicit stakeholders’ opinions on the current status and impact of health sciences libraries and their own information needs. A workshop and planning session was convened in Morogoro, followed by a series of four focus groups, to gather input from different communities in sites across the country. Input from a variety of interested individuals—physicians, nurses, librarians, trainers, academicians, students, patients, and families—was gathered.

A second conference, with a video connection to NLM, was held in Moshi, in October 2014. Meeting participants included members of NAML, senior medical librarians from Duke University, Johns Hopkins University, University of Maryland-Baltimore, University of Michigan, University of North Carolina, University of Pittsburgh, and Stanford University, as well as NLM staff from LO and the National Center for Biotechnology Information. NLM’s Office of High Performance Computing and Communications in the Lister Hill National Center for Biomedical Communications provided technical support for the videoconference, along with the ICT department at Kilimanjaro Christian Medical College.

During the two workshops, stakeholders identified key weaknesses of current medical libraries in Tanzania and began developing a three-year Diploma program to address them. Conference participants continue to collaborate on this project.

Scientific Capacity Building

OHIPD provides support for education and training activities in biomedical informatics for African researchers. In FY2014, training was provided for two workshops with scientists from several African countries. These were *Finding, Organizing, and Using Medical/Scientific Information*, Nador, Morocco, October 22-23, 2014; and *Workshop on Newborn Screening in Morocco, Rabat, Morocco*, December 9-11, 2014.

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 NLM 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 mHealth Working Group and the NIH MEPI working group.

International Visitors

In FY2014, 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 42 countries: Albania, Argentina, Bangladesh, Bosnia and Hercegovina, Canada, China, India, Iran, Israel, Japan, Jordan, Korea, Mexico, Morocco, Netherlands, Nigeria, Norway, Pakistan, Russia, Saudi Arabia, Slovenia, South Korea, Sri Lanka, and Sweden.

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 FY2014, 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 NLM Library Operations (LO) Division is responsible for ensuring access to the published record of the biomedical sciences and the health professions. LO acquires, organizes, and preserves NLM's comprehensive archival collection of biomedical literature; creates and disseminates controlled vocabularies and a library classification scheme; produces authoritative indexing and cataloging records; builds and distributes bibliographic, directory, and full-text databases; provides national backup document delivery, reference service, and research assistance; helps people to make effective use of NLM products and services; and coordinates the National Network of Libraries of Medicine to equalize access to health information across the United States. These essential services support NLM's outreach to health professionals, patients, families and the general public, as well as focused programs in AIDS information, molecular biology, health services research, public health, toxicology, environmental health, and disaster planning.

Library Operations also develops and mounts historical exhibitions; produces and manages a travelling exhibition program; creates and promotes education and career resources for K-12 and undergraduate students and educators; carries out an active research program in the history of medicine and public health; collaborates with other NLM program areas to develop, enhance, and publicize NLM products and services; conducts research related to current operations; directs and supports training and recruiting programs for health sciences librarians; and manages the development and dissemination of national health data terminology standards. LO staff members participate actively in efforts to improve the quality of work life at NLM, including the work of the NLM Diversity Council.

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

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

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

Program Planning and Management

LO sets priorities based on the goals and objectives in the NLM Long Range Plan 2006-2016, and the closely related NLM Strategic Plan to Reduce Racial and Ethnic Disparities. In FY2014, LO continued its work on the directions of its Strategic Plan for 2010-2015, within this broader framework.

The Technical Services Division (TSD) officially reorganized as of January 26, 2014. All selection, acquisitions, and licensing functions were integrated in one section, the Collection Development and Acquisitions Section. The Cataloging Section changed its name to the Cataloging and Metadata Management Section. The reorganization created a new section, the Library Technology Services Section, to manage the Division's systems support functions.

In the area of Developing a 21st Century Workforce, LO held four quarterly all-staff meetings in which new staff are recognized and updates from every area of LO are presented to those in the auditorium as well as staff listening in from off-site. LO continued its second full year of a Career Enrichment Program, a professional development program for selected LO staff to obtain a broader view of LO and NLM and to work on a project of institutional significance. LO also held ongoing discussion groups for supervisors. Capitalizing on a new Federal hiring initiative, Pathways for Recent Graduates, LO hired 17 recent library science graduates, representing 6 percent of its Federal workforce, in three Divisions. The new employees will work in acquisitions, e-resource management, reference, preservation, outreach, digitization, digital preservation, Web development, social media analysis and deployment, systems, data analysis, customer services, Web user experience, and education and training.

Collection Development and Management

The NLM comprehensive collection of biomedical literature is the foundation for many of the Library's services. LO ensures that this collection meets the needs of current and future users by updating the NLM 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 FY2014, the NLM collection contained 2,781,201 volumes and 23,867,056 other physical items, including manuscripts, microforms, prints, photographs, audiovisuals, and electronic media.

Selection

Publishing trends had an impact on the selection of new journals for the collection. The number of newly-launched journals decreased at some major medical publishers. The trend toward the “mega-journal” (a single journal that covers a wide number of disciplines) had an effect, as several publishers entered that arena and offered what might have previously constituted many new journals as a single title. Selection activity increased due to the adoption of somewhat more rigorous journal selection guidelines, necessitated in part by the proliferation of journals that do not meet the most basic standards of medical publishing.

Following the discovery of a large collection of uncatalogued World Health Organization documents that had been given to NLM in the past, selectors reviewed and sent many titles for cataloging. This collection includes reports on malaria and other infectious diseases, brief papers by notable medical scientists such as Jonas Salk, and reports on health conditions in Africa and Latin America, primarily from the 1940s and 1950s. The review of this collection will continue into the new fiscal year.

In their efforts to enrich the NLM collections, selectors focused on areas of critical national and international importance. For example, in response to the Ebola crisis, selectors identified books, reports, and video recordings on the science and history of the disease, as well as preparedness and response documents. HMD and TSD staff also worked to launch a Web collecting initiative to capture and preserve selected born-digital content documenting the Ebola outbreak. Examined content included Web sites and social media from Government and non-government organizations, journalists, healthcare workers, and scientists in the United States and around the world, with an aim to collect and preserve a diversity of perspectives on this health crisis.

Web content on other infectious diseases (such as influenza and tuberculosis) and topics such as health care reform, global health, and environmental health disasters were also acquired. The collecting rationale is to assemble a collection of works that are of interest to current researchers and that also chronicle health-related events that will be of interest to researchers in the future. National Digital Stewardship Resident Maureen Harlow conducted a project to collect Web content on Disorders of the Developing and Aging Brain: Autism and Alzheimer’s. LO also continued to collect blogs authored by doctors and patients, to illuminate health care thought and practice in the 21st century.

Acquisitions

TSD received and processed 114,197 contemporary physical items (books, serial issues, audiovisuals, and electronic media). The number of electronic-only serials grew to nearly 3,200 by the end of FY2014, now representing more than 18 percent of all currently acquired serials. In FY2014, 5,547 licensed and 4,415 free

electronic journals were available to NLM users. A net total of 28,911 volumes and 4,810,440 other items (including non-print media, manuscripts, and pictures acquired by HMD) were added to the NLM collection.

Late in September 2014, NLM learned that Swets Information Services, a company that served as the primary serials subscription agent for NLM, filed for bankruptcy. The company based in the Netherlands, provided subscription services for hundreds of libraries around the world. Swets managed NLM orders for approximately 8,000 serial titles from over 3,300 different publishers in 66 countries. NLM was able to de-obligate the balance of funds from the contract prior to the end of FY2014, which prevented any loss of funds as a result of the bankruptcy. Orders for 2015 subscriptions will be handled by a new contractor (or contractors). The shut-down of Swets’ facilities interrupted the delivery of issues to the Library, resulting in short-term gaps in the print collection. Many publishers contacted NLM and offered to ship issues directly, and staff contacted other publishers to request that issues be mailed to NLM. Electronic access was not impacted. Over 60 percent of the titles subscribed to via Swets are available electronically.

HMD acquired a wide variety of important printed books, manuscripts and modern archives, images, and historical films during FY2014, including an early German manuscript pharmacopoeia by an anonymous author, written around the year 1600, with later additions included. The pharmacopoeia is a recipe book with formulae for waters, electuaries, oils, ointments, etc., for combatting cancer, plague, jaundice, fevers, kidney and liver ailments, gynecological disorders, burns, fractures and other infirmities. Following the main text is a circa 1800 list of common abbreviations, an unfinished glossary (going only to ‘C’) of Latin chemicals and ingredients with German translations and a brief note about each, and an alphabetical list of ailments and conditions listing the main ingredients to be found in medicaments for their treatment.

Among the important printed books acquired in FY2014 are several early foreign language editions of Charles Darwin’s *On the Origin of Species by Means of Natural Selection*, including versions in Russian, French, and Hungarian. The foreign translations of this groundbreaking work are important because they often include unique commentary by the translators, and because the theory of evolution that was promulgated in the book was received differently all over the world. These books provide insights into how the theory was presented, debated, and quickly spread throughout the scientific world.

NLM received a large collection of AIDS-related books as a gift from Dr. June E. Osborn. During the 1980s and 1990s, Dr. Osborn held numerous senior positions, including Chair of the National Institutes of Health National Heart, Lung and Blood Institute advisory committee on AIDS, the National Advisory Committee for the Robert Wood Johnson Foundation’s AIDS Health

Services Project, and the US National Commission on AIDS. She was also a member of the Global Commission on AIDS of the World Health Organization.

The Library received a large donation of materials, many in Russian, from the National Aeronautics Space Administration (NASA). Many of the titles were included in the SPACELINE subset of MEDLINE but not previously held by NLM. (Between 1993-2005, information about space life sciences was provided by NLM in conjunction with the NASA SPACELINE Office as part of a collaborative agreement). This gift has enriched the NLM collection of works on space life sciences.

Significant acquisitions in the Archives and Modern Manuscripts Program during FY2014 included the electronic files of former Surgeon General Regina Benjamin, the archival records from HealthNet News/Satellife, and 67 boxes of archival materials from former Surgeon General C. Everett Koop's family, which will be added to the existing NLM Koop collection. The Historical Audiovisuals program received a donation of 25 surgical training films produced by Davis & Geck of Danbury, Connecticut. This donation complements and expands our existing collection of Davis & Geck films.

Preservation and Collection Management

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

Collection Space and Maintenance

In FY2014, LO bound 14,516 volumes, repaired 685 items, made 811 preservation copies of films and audiovisuals, and conserved 583 items. A total of 363,277 items were shelved, a 9 percent decrease from FY2013 that reflects the decline in print material received at the Library and the decline in interlibrary loan and in requests by patrons in the Main Reading Room, due to increased availability of electronic journals.

Work continued on the long term project to install compact shelving on the B-2 level to increase storage capacity for collections in the NLM building. The project requires floor strengthening on the B-2 level and includes the complete upgrade of the fire suppression system and lighting for stack areas on the B-2 and B-3 levels, as well as an upgrade of the older compact shelving on the B-3 level. New space was created for the Preservation and Collection Management (PCM) Section staff, and future plans call for new space for HMD staff on the B-1 and B-2 levels. At the end of 2014, the project is 60 percent complete, providing collections growth space until 2022.

When the project is done, collections can grow until at least 2030.

National Cooperative Preservation

MedPrint is the National Network of Libraries of Medicine (NN/LM) cooperative project to preserve key biomedical journals in print until there is stronger evidence for the reliability of digital preservation. The program is open to all US libraries that participate in DOCLINE, the NLM interlibrary loan (ILL) system that stores journal holdings information for almost 2,500 libraries. Print retention commitments are also stored in DOCLINE. By the end of 2014, NLM had 21 signed agreements from institutions representing all eight regions. One hundred and one libraries have recorded print retention commitments for 1,376 titles.

Digitization Program

The NLM Digital Collections repository now holds over 12,000 monographs and serials and over 150 films, with newly digitized texts and films added regularly. NLM Digital Collections also ingested the approximately 3.8 million citations from the IndexCat Web site, making the contents of the Index Catalogue of the Surgeon General's Library more readily available. During the year, Digital Collections was modified to support ingest of serial publications, and the Web site was enhanced to provide a hierarchical display of each serial's digitized holdings.

The installation of CCS docWorks (dW) image processing software was a major enhancement to LO's digitization program, providing a more efficient scanning workflow, greater capabilities to crop and de-skew images, and the ability to analyze the structure and content of the digital surrogates resulting in enhanced structural metadata files for the digital books.

Preservation and Collection Management and History of Medicine staff completed digitization for several projects including an important collection of World War I 1914-1922 monographs. Combined with the ongoing digitization initiatives, including Medicine in the Americas Phase 2, NLM Publications, and the Scan on Demand program for interlibrary loan (ILL), a total of 2,284 volumes and 599,934 pages were digitized in FY2014.

PubMed Central (PMC), a digital archive of medical and life sciences journal literature developed by NCBI, is the NLM vehicle for ensuring permanent access to electronic journals and digitized back files. LO assists NCBI in soliciting participation of additional journals, particularly in the fields of clinical medicine, health policy, health services research, and public health. LO provides support for the PMC efforts ranging from review of potential journals for appropriateness for the NLM collection to cataloging and authority data creation for the PMC system. By the end of FY2014, 207 new journals had been added to PMC, and 368,111 new articles had been

added. The total number of articles in the database is 3,227,379.

In April 2014, NLM and the Wellcome Trust signed a Memorandum of Understanding to work together to make thousands of complete back issues of historically-significant biomedical journals freely available in PMC. Building on the NLM and Wellcome Trust's previous Medical Journal Backfiles Digitization Project of 2004-2010, the terms of the new agreement include a donation of £750,000 (\$1.2 million) to NLM to support coordination of the digitization project, which will be done using NLM materials. Part of the project will concentrate on mental health journals, supporting a major archive digitization program being undertaken by the Wellcome Trust. The project will add between 1.2 to 5 million additional scanned pages to PMC over the course of three to five years. In 2014, the Web Collecting and Archiving Working Group began exploring the use of Web crawling technologies to archive NLM Web sites and replace the current content management system used for this function.

Vocabulary Development and Standards

LO produces and maintains the Medical Subject Headings (MeSH), a subject thesaurus used by NLM and many other institutions to describe the subject content of biomedical literature and other types of information; develops, supports, or licenses for US use vocabularies designated as US standards for electronic health records; and works with OCCS to produce the Unified Medical Language System (UMLS) Metathesaurus, a large vocabulary database that includes more than 100 vocabularies, including MeSH and several others developed or supported by NLM. The Metathesaurus is a multi-purpose knowledge source used by NLM and many other organizations in production systems and informatics research. It serves as a common distribution vehicle for classifications, code sets, and vocabularies designated as standards for US health data.

Medical Subject Headings (MeSH)

Work was completed on the 2015 edition of MeSH which contains 27,455 descriptors, 224,925 supplementary records for chemicals, other substances and diseases, and 83 qualifiers. The MeSH Section added 310 new descriptors, replaced 25 descriptors with more up-to-date terminology, and deleted 4 descriptors.

MeSH worked with the Medical Language Branch, OCCS to develop functional requirements for a new MeDit editing system for the MeSH vocabulary to replace M2000 which had been in use since 1999. MeSH tested successive versions of the new software for needed functionality and after extensive testing and revision, the MeDit software was released for this production year. MeSH staff developed training materials and led several hands-on training sessions for Index Section chemists and other LO members who need access to MeDit software.

MeSH staff developed a new SharePoint site that serves production management needs of our yearly release cycle. The new site provides an NLM-wide entry point for library employees to enter MeSH requests. The vocabulary requests are then tracked by SharePoint as they are acted on by MeSH analysts and work their way through our internal and external review cycles. The site provides a way for external participants to enter comments and to collaborate on a variety of topics.

A new service, MeSH on Demand, finds MeSH terms relevant to input text up to 10,000 characters (such as manuscript abstract, summary statement or description of a project) automatically. In the first full month of operation, 140,940 requests were made to MeSH on Demand, which was developed by LO and the Lister Hill National Center for Biomedical Communications.

In FY2014, LO introduced a new workflow for proposing and reviewing changes to MeSH qualifiers (subheadings). Two pilot proposals from BSD and TSD were submitted, reviewed, and approved. Internal MeSH production systems were reviewed for future changes in order to accommodate these proposals.

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

UMLS Metathesaurus

Using systems managed by OCCS, the MeSH Section acquires, inserts, and edits updates to the vocabularies included in the UMLS Metathesaurus and manages the twice yearly production schedule. BSD manages quality assurance procedures and customer service for UMLS users, including licensing, documentation, and educational materials. CVX (vaccines administered from the CDC) was added as a new source in the November 2013 release of the UMLS Metathesaurus.

Clinical Vocabularies

NLM is the central coordinating body for clinical terminology standards within the Department of Health and Human Services (HHS). LO, in partnership with LHCBC and OCCS, represents NLM in Federal initiatives to select and promote use of standard clinical vocabularies in electronic health records as well as administrative transactions governed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). With enactment of the Health Information Technology for Economic and Clinical Health (HITECH) Act, included as part of the American Recovery and Reinvestment Act of 2009 (ARRA), the NLM activities in

this area intensified, particularly in the areas of quality and performance measurement, lab services, genetics, and newborn screening. FY2014 saw continued efforts by NLM to promote use of health data standards both in research and more broadly.

With support from an interagency agreement (IAA) with the Centers for Medicare and Medicaid Services (CMS) initiated in FY2012, NLM developed the Value Set Authority Center (VSAC), in conjunction with both CMS and the Office of the National Coordinator for Health Information Technology (ONC).

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

NLM also has an interagency agreement with the Veterans Health Administration (VHA) designed to accelerate the pace of clinical terminology standards development and integration in areas that support Veterans health care and benefits determination. Areas covered by this agreement include alignment of SNOMED CT with other terminologies designated for use in national healthcare systems (e.g., LOINC and RxNorm) as well as enhancing these standards as appropriate to improve coverage for recording diet and nutrition, functional status, mental health, as well as medical devices.

In FY2014, NLM continued work related to the identification of national standards for description and use of structured data from electronic health records (EHRs) and the development of a NLM/NIH (National Institutes of Health) common data element repository. Supported by the Office of the Secretary's Patient Centered Outcomes Research Trust Fund for data infrastructure, the structured data capture project is intended to facilitate the use of data captured by EHRs for comparative effectiveness research (CER), public health, adverse event and patient safety reporting, and other purposes. The project complements efforts led by NLM in conjunction with the trans-NIH Biomedical Informatics Coordinating Committee (BMIC) to promote data harmonization, quality, and reuse in clinical research. In FY2014, a working group of BMIC expanded information available on the NIH Common Data Element (CDE) Resource Portal. Through identification of NIH research initiatives that use CDEs, the Resource Portal and working group encourage use of common data elements in clinical research, patient registries, and other human subject research in the NIH community and beyond.

RxNorm

The MeSH Section produces RxNorm, a clinical drug vocabulary that provides standardized names for use in prescribing. It is released through the UMLS and also separately. RxNorm represents the information that is typically known when a drug is prescribed, rather than the specific product and packaging details that are available at the time a medication is dispensed or administered, and provides a mechanism for connecting information from different commercial drug information services. RxNorm editors prepare and release weekly additions and full monthly updates to the clinical drug vocabulary.

In addition to reflecting new drugs as approved by the FDA, in FY2014 RxNorm content was enhanced to increase its utility in physician ordering applications, to link it more effectively to FDA's unique ingredient identifiers (UNII), to label obsolete drug names and National Drug Codes more accurately, and to accommodate a broader range of over the counter drugs. All of this improved RxNorm content, and was done in concert with RxNorm source providers such as First DataBank, Multum, and the (US Department of Veterans Affairs (VA).

LOINC

Through LO's NICHSR, NLM supports the continued development and free distribution of LOINC® (Logical Observation Identifiers Names and Codes) by the Regenstrief Institute. LOINC is a clinical terminology important for laboratory test orders and results. The final rule for 2014 EHR certification criteria, published in September 2012 and updated in 2014, expanded the requirements for use of LOINC. LOINC was also listed as a requirement under the proposed rule for 2015 EHR certification criteria published in February 2014. In FY2014, NLM established a new five-year follow-on contract with the Regenstrief Institute to ensure the continued ongoing maintenance and free distribution of LOINC. Additionally, NLM staff, primarily in LHCNBC, also made substantial contributions to enhancing its content.

SNOMED CT

In FY2014, NLM continued to serve as the US Member of the International Health Terminology Organization (IHTSDO) and to pay the annual fees for the US-wide license for SNOMED CT, the comprehensive clinical terminology which the IHTSDO owns. The final rule for 2014 EHR certification criteria published in September 2012 and updated in 2014 expanded the requirements for use of SNOMED CT. SNOMED was also listed as a requirement under the proposed rule for 2015 EHR certification criteria published in February 2014. NLM is responsible for distribution of SNOMED CT within the US and for representing US interests in the continued

development of SNOMED CT. The IHTSDO continues to grow and now includes 27 Member countries including the major English-speaking countries.

FY2014 saw several important advances that will improve the use of SNOMED CT in the US and elsewhere. NLM continued maintaining both the US SNOMED CT Content Request Submission System as well as the SNOMED CT International Request Submission System (SIRS), the second on behalf of the IHTSDO. NLM continued distributing twice annual updates of the US Extension to SNOMED CT. Because the US Extension to SNOMED CT is intended to be used with the International Release of SNOMED CT, the two were combined into the US Edition of SNOMED CT which is now distributed twice a year. In 2014, the IHTSDO published new content generated through various cooperative agreements, including the LOINC-SNOMED Cooperation Project Technology Preview, the Candidate Baseline of General Practice/Family Practice Reference Sets and SNOMED CT to International Classification of Primary Care Maps, and the SNOMED CT International Medical Devices release.

NLM continues to be involved in activities related to 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 by connecting them to billing systems, with the ultimate goal of helping to improve the overall quality and efficiency of health care services. In 2014 the two Memoranda of Understanding, one between the IHTSDO and Kaiser and one between NLM and Kaiser, relating to Kaiser's Convergent Medical Terminology (CMT) were renewed. Under these agreements, the Kaiser contributions included subsets for common lab procedures, emergency department, injuries, orthopedics, ophthalmology, musculoskeletal conditions, cardiology, and pediatrics. Where appropriate, CMT content is incorporated into the International Release of SNOMED CT or the US Extension to SNOMED CT when only relevant for use in the US.

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.

Cataloging

LO catalogs the biomedical literature acquired by NLM to document what is available in the Library's collection and to identify high-quality, relevant medical resources on the Web. NLM also works with the Library of Congress (LC) to provide cataloging-in-publication (CIP) data for US medical books prior to their publication. Libraries worldwide use the cataloging and authority data created by NLM, thus reducing local cataloging effort.

TSD is responsible for maintaining and updating the NLM Classification and the NLM Catalog Document Type Definition for transforming Machine-Readable Cataloging (MARC) data into other metadata formats and vice-versa, and for name authority control. The 2014 edition of the NLM Classification was released in online and PDF formats providing updates to the QS (Human Anatomy) and QT (Physiology) schedules.

TSD participated in the BIBFRAME Phase 2 implementation group assembled by the Library of Congress. 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. NLM tested, compared and provided feedback on transformation and input tools provided by the Library of Congress and Zepheira.

In FY2014, TSD cataloged a total of 18,775 books, serials, electronic resources, and audiovisuals. This figure includes 2,156 Cataloging-in-Publication (CIP) records created from publisher electronic galleys or ONIX data and 2,155 titles cataloged for the HMD collection.

Indexing

LO indexes 5,647 biomedical journals for the MEDLINE database to assist users in identifying articles on specific biomedical topics. A combination of Index Section staff, contractors, and cooperating US and international institutions indexed 765,850 articles in FY2014, bringing the total number of MEDLINE citations to over 21.7 million. Indexed citations were updated to reflect 500 retracted articles, 9,602 published corrections, and 56,652 comments found in subsequently published notices or articles.

In FY2014, indexers created 78,491 annotated links between newly indexed MEDLINE citations for articles describing gene function in selected organisms and corresponding gene records in the NCBI Entrez Gene database. This was an 8 percent decrease from last year.

LO continues to work with other NLM program areas to identify, test, and implement ways to reduce or eliminate tasks now performed by human indexers. At the end of FY2014, 4,376 journals (77 percent of all indexed journals) were indexed from an online version, including online-only journals and those that also have a print version.

The Medical Text Indexer First Line (MTIFL) workflow became a regular part of work for all in-house indexers. During FY2014, 63,456 articles from 211 journals were indexed using MTIFL, which is a combination of machine-suggested indexing which is then reviewed and approved by senior indexers. Expanding the use of MTIFL is a Government Performance and Results Act (GPRA) goal for NLM. The goal for FY2014 was to have 186 journals added to MTIFL; NLM exceeded this goal by 25 journals.

Indexers perform their work after the initial data entry of citations and abstracts has been accomplished. Over the past 12 years, great strides have been made in improving the efficiency of data entry. By the end of FY2014, 94 percent or 926,561 citations were submitted in XML format by publishers, a seven percent increase over the previous year. The 55,222 remaining citations were created by scanning and optical character recognition (OCR), for a total of 981,783 citations created in FY2014.

NLM selects journals for indexing with the advice of the Literature Selection Technical Review Committee (LSTRC) (**Appendix 11**), an NIH-chartered committee of outside experts. In FY2014, LSTRC reviewed 493 journals and recommended 73 for MEDLINE indexing.

Information Products

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

Databases

LO manages the creation, quality assurance, and maintenance of the content of: MEDLINE/PubMed, the NLM database of indexed journal citations; the NLM catalog; MedlinePlus and MedlinePlus en español, NLM's primary health information resources for the general public; and a number of specialized databases, including several in the fields of health services research, public health, and history of medicine. These databases are richly interlinked with each other and with other important NLM resources, including PMC (PubMed Central); scientific data resources; ClinicalTrials.gov; Genetics Home Reference; and toxicological, environmental health, and AIDS information services produced by NLM's Specialized Information Services Division.

Use of MedlinePlus increased to 409 million unique visitors in FY2014, a 21 percent increase from FY2013. MedlinePlus en español saw a 37 percent increase, with 298.5 million unique visitors in FY2014. Use of MedlinePlus Mobile increased 37 percent to 1,481,000 unique visitors. Requests to MedlinePlus Connect increased to 67.8 million in FY2014 from 25.5 million in FY2013. MedlinePlus, MedlinePlus en español and MedlinePlus Mobile continue to receive high ratings from customers in the American Customer Satisfaction Index (ACSI), ranking among the top government news/information sites.

PSD and OCCS continued to expand and improve the suite of MedlinePlus-branded products and services. At the end of FY2014, MedlinePlus featured 957 topics in English and 948 in Spanish. Three major MedlinePlus Connect enhancements contributed to the large increase in

use of that product: support of ICD-10-CM (International Classification of Diseases, 10th edition, Clinical Modification), the integration of information from Genetics Home Reference for a subset of MedlinePlus Connect responses, and enhancing the Web service response to offer JSON and JSONP formats as well as an updated XML format that matches the latest HL7 (Health Level 7 International) standard.

At the end of FY2014 the English and Spanish Twitter feeds had 134,376 followers. Subscriptions to the MedlinePlus and MedlinePlus en español GovDelivery email update service grew to 729,277 subscribers with 72.4 million subscriptions, up from 623,651 subscribers with 65.1 million subscriptions at the end of FY2013.

Under the leadership of NICHSR, NLM continues to expand and enhance its databases for health services research and public health. NICHSR continues to find new sources of grey literature for the NLM collection, providing timely access to new and emerging research and analyses, essential to the needs of policy and health care professionals. In FY2014, 18 new sources were added to the set routinely reviewed for HSR-relevant topics; 436 items were added to the NLM Catalog and 318 citations to articles added to MEDLINE. NICHSR collaborates with NCBI to maintain and expand HSR, public health, and clinical guidelines-related materials available on Bookshelf.

In FY2014, NICHSR continued to improve the depth and breadth of content on its information portal for the health services research community, HSR Information Central (HSRIC). A new topic page "Community Benefit/Community Health Needs Assessment" was added to the site, highlighting resources related to fulfillment of recent Internal Revenue Service (IRS) requirements for hospital not-for-profit status. In FY2014 page visits increased by 10 percent and page views increased by over 15 percent. In FY2014, HSRIC joined the GovDelivery communications platform, increasing the visibility of its pages and enabling users to receive customized updates about additions to the site and its topic pages.

NICHSR contributes to the field of health services research through databases containing information about newly-funded research as well as accessible data collection tools and resources useful to researchers. HSRProj (Health Services Research Projects in Progress), managed by Academy Health and the Sheps Center at the University of North Carolina-Chapel Hill, contains HSR projects funded by government, foundations, and other organizations, such as the Patient Centered Outcomes Research Institute. In FY2014, the database grew to more than 12,000 active records representing more than 150 funders. It is frequently used to assess both research gaps and trends. For example, in FY2014, AcademyHealth and the Association of American Medical Colleges used HSRProj to examine national trends in health disparities research topics and investments. Use of the database increased dramatically in FY2014 – the number of visitors more than doubled, page views increased more than 30

percent and visits increased almost 70 percent. In FY2014, the database Health Services Research Resources (HSRR) also expanded, now covering over 1500 datasets, surveys, other research instruments, and software packages used with datasets.

DailyMed, a Web site that presents high quality information about drugs, including the FDA approved packaging information (labels) for drugs, grew tremendously throughout the year to approximately 20,000 labels; 75-100 updates are received each day. Over the past year grandfathered prescription drugs and the over-the-counter product labels were also added. Use of DailyMed was 17,967,944 visitors in FY2014, a 9 percent increase over FY2013.

In September 2014 NLM sponsored an SPL/DailyMed Jamboree Workshop – Practical use of DailyMed and RxNorm Drug Data. Speakers from Federal government, industry, academia and non-profit sectors spoke on their experience with Structured Product Labeling (SPL) drug data as well as RxNorm. The emphasis was on practical and novel ways to use this free data, which is produced cooperatively by NLM and FDA. The Jamboree was a big success with 202 registrants.

Machine-Readable Data

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

Demand for MEDLINE/PubMed data in XML format continues to increase. At the end of FY2014, there were 831 licensees of MEDLINE data, a 9 percent increase over the previous year. The majority use the data for research and data mining.

At the end of FY2014, there were 12,627 UMLS licensees, an increase of 43 percent from the previous year. NLM services and support to licensees of UMLS were enhanced by user input received in UMLS Annual Reports licensees are required to submit.

Web and Print Publications

NLM databases and Web sites are its primary publication media. Publications available on the main Web site include recurring newsletters and bulletins, fact sheets, technical reports, and documentation for NLM databases. The BSD MEDLARS Management Section edits and produces the *NLM Technical Bulletin*, which provides timely, detailed information about changes and additions to the NLM databases and related policies, primarily for librarians and other information professionals. Published since 1969, the

NLM Technical Bulletin also serves as the historical record of the evolution of the NLM online systems and databases. In FY2014, LO staff continued to be involved in the two publications designed for patients, families, and the public. The Director's Comments podcasts bring current health news to listeners. Four issues of the *NIH MedlinePlus Magazine* and one issue of *NIH MedlinePlus Salud* were published in print and online in FY2014.

Direct User Services

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

Reference and Customer Services

LO provides reference and research assistance to onsite and remote users as a backup to services available from other health sciences and public libraries. LO also has primary responsibility for responding to inquiries about the NLM products and services and how to use them effectively. PSD responds to initial inquiries and also handles the majority of questions requiring second-level attention. Staff from throughout LO and NLM assist with second-level service when their special expertise is required. LO Customer Services received 118,158 inquiries in FY2014. PSD combined the separate Reference and Circulation desks in the Main Reading Room into one Information desk in FY2014, which enabled more accurate counting of onsite inquiries. As a result, onsite inquiries counted increased 33 percent to 6,967. The number of remote inquiries (95,665) decreased 4 percent.

Document Delivery

LO retrieves documents requested by onsite patrons from the NLM closed stacks and also provides interlibrary loan as a backup to document delivery services available from other libraries and information suppliers. In FY2014, PSD processed 257,193 requests for contemporary documents, a 13 percent decrease. The History of Medicine Division (HMD) handled 8,660 requests for rare books, manuscripts, pictures, and historical audiovisuals, a 13 percent increase.

Onsite

The number of requests by users in the Main Reading Room to the NLM closed-stack collections declined by 10 percent to 68,281 requests. Many users take advantage of access to online journal subscriptions in the Main Reading Room, eliminating the need for stack requests.

OffSite

The Collection Access Section (CAS) received 188,912 interlibrary loan requests, a 16 percent decline from FY2013, and filled 83 percent of them. The decline in ILL requests to NLM was due in part to the Government-wide shut down for three weeks in October 2013, but also to a general decrease in demand for resource sharing among NN/LM members. NLM ILL staff keyboarded a total of 42,283 requests, a 3 percent increase from FY2013, many from international libraries. In FY2014, 94 percent of requests were filled within 4 hours of receipt and 98 percent were filled within one business day of receipt. NLM now delivers 97 percent of interlibrary loan requests electronically.

The Strategic Planning Working Group formed to study DOCLINE and Resource Sharing in the National Network of Libraries of Medicine (NN/LM) completed its work in FY2014. The Group had conducted three focus groups with network members and an online survey of DOCLINE members. Its report identified reasons for the decline in DOCLINE use: 1) the increased availability of electronic journal literature via public access initiatives and open access; 2) changing end-user expectations and behavior to primarily select items available immediately rather than request ILL; and 3) the increased use of synthesized point-of-care information tools in lieu of journal articles. The team recommended implementing some DOCLINE system changes to improve overall system efficiency, and to allow network members greater flexibility in how often and at what level they participate in DOCLINE.

A total of 2,497 libraries use DOCLINE, the NLM interlibrary loan request and routing system, a decrease of 2 percent from FY2013. DOCLINE users entered 1,280,726 requests in FY2014, a 9 percent decline from last year. There were 218,944 Loansome Doc requests entered into the system by PubMed users, down 12 percent from FY2013.

There are 1,712,556 holdings records in DOCLINE, of which 1,059,341 are for print journals; 626,117 for electronic; and 27,098 for "other." The electronic holdings now constitute over 36 percent of holdings, up from 33 percent in FY2013.

NCBI, BSD, and the staff at the Regional Medical Libraries continued to support and promote the use of the PubMed LinkOut for Libraries and Outside Tool, the open-URL services that allow libraries to link directly from PubMed to a wide range of resources beyond the Entrez system. Using these tools, libraries can create custom displays of their electronic and print holdings for their primary clientele. The number of libraries participating in LinkOut increased by 6 percent in FY2014 to 3,127; there are 1,190 libraries participating in the Outside Tool option, an increase of 10 percent over last year.

NLM and the Regional Medical Libraries continued to encourage network libraries to use the Electronic Funds Transfer System (EFTS), which was

developed to reduce administrative costs associated with interlibrary loan service billing and is operated for the NN/LM by the University of Connecticut on a cost recovery basis. At the end of FY2014, there were 1,395 libraries participating in EFTS. The decline in the number of libraries and the number of transactions in DOCLINE has caused a similar decline in membership and transactions in EFTS.

Outreach and Training

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

National Network of Libraries of Medicine (NN/LM)

The NN/LM works to provide timely, convenient access to biomedical and health information for US health professionals, researchers, and the general public irrespective of their geographic location. With over 6,300 full and affiliate members, the Network is the core component of the NLM outreach program and its efforts to reduce health disparities and to improve health information literacy. Full members are libraries with health sciences collections, primarily in hospitals and academic medical centers. Affiliate members include some smaller hospitals, public libraries, and community organizations that provide health information service, but have little or no collection of health sciences literature. LO's National Network Office (NNO) oversees network programs that are administered by eight Regional Medical Libraries (RMLs) under contract to NLM.

RMLs and other Network members conduct many special projects to reach underserved health professionals and to improve the public's access to high quality health information. Virtually all of these projects involve partnerships between health sciences libraries and other organizations, including public libraries, public health departments, professional associations, schools, churches, and other community-based groups. In FY2014, the NN/LM initiated 92 outreach projects which target rural and inner city communities and special populations in 28

states, the District of Columbia, US Virgin Islands, Puerto Rico, and Guam.

With the assistance of other NN/LM members, the RMLs staff most of the exhibits and demonstrations of NLM products and services at health professional, consumer health, and general library association meetings around the country. LO organizes the exhibits at the Medical Library Association annual meeting, and, 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 FY2014, NLM and NN/LM services were exhibited at 23 national and 377 regional, state, and local conferences across the US. These exhibits highlight all NLM services relevant to attendees.

Partners in Information Access for the Public Health Workforce

The NN/LM is a key member of the Partners in Information Access for the Public Health Workforce, a 14-member public-private agency collaboration initiated by NLM, the Centers for Disease Control and Prevention, and the NN/LM in 1997 to help the public health workforce make effective use of electronic information sources and to equip health sciences librarians to provide better service to the public health community. The NICHSR coordinates the Partners for NLM; staff members from the National Network Office 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 at the University of Massachusetts, provides unified access to public health information resources produced by all members of the Partnership, as well as other reputable organizations. In FY2014, the Web site was expanded with more than 587 new links and was responsively redesigned to enhance access from any kind of mobile device, in addition to desktop computers. The site also joined the GovDelivery communications platform in FY2014. From FY2013 to FY2014, visitors and visits increased about 8 percent, while page views increased almost 50 percent.

Among the most popular resources on the Partners Web site are pre-formulated PubMed search queries in support of the HHS Healthy People initiative. NICHSR has worked with the HHS Office of Disease Prevention and Health Promotion, responsible for coordinating the development and management of the Healthy People initiative, to arrange for the development of structured searches for Healthy People 2020, which launched at the end of 2010. In FY2014, page views of the Structured Evidence Queries (SEQs) from the Partners Web site increased by 33 percent. The SEQs were presented in a poster at the annual meeting of the American Public Health Association. The SEQs are accessible both from the PHPartners.org site and from the HealthyPeople 2020 site, at www.healthypeople.gov.

Special NLM Outreach Initiatives

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

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

Historical Exhibitions and Programs

HMD'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 resources, foster partnerships with other Federal agencies and cultural organizations, and are an important part of the NLM outreach program. The Library's flagship exhibition on display in FY2014, located in the Rotunda, continued to be *Native Voices: Native Peoples' Concepts of Health and Illness*, which explored the interconnectedness of wellness, illness, and cultural life for Native Americans, Alaska Natives, and Native Hawaiians. Through the extensive use of interviews and media, visitors discover how Native concepts of health and illness are closely tied to the concepts of community, spirit, and the land.

During FY2014, NLM traveled 30 copies of 13 exhibition titles sending shows to 198 host venues, 113 of which were members of the National Network of Libraries of Medicine (NN/LM). NLM exhibitions traveled to 44 states and three international locations. Host venues reported 674,358 people visited an NLM traveling exhibition during FY2014. The hosts sponsored 80 public programs related to NLM exhibitions.

HMD featured special displays in its reading room, complemented by dynamic online versions that reach a wide public audience. The special display and online exhibition *From DNA to Beer: Harnessing Nature for Science & Industry*, a collaborative effort with the National Museum of American History, opened to the public November 5, 2013. The special display and online exhibition *Pictures of Nursing: the Zwerdling Postcard Collection* opened to the public September 2, 2014. Events included a lecture by the guest curator Julia Hallam, PhD followed by a reception and an invitational tour of a local area nursing college.

HMD develops resources to support classroom educators and the engagement of students with online exhibitions to introduce these audiences to NLM collections and resources. HMD hosted the third annual

Education Institute June 15-27, 2014. Seven public school librarians from three area school systems (District of Columbia, Montgomery, and Prince George's counties) participated in the Institute, during which they provided feedback on existing exhibition educational resources, traveling exhibitions, and building curricular connections to the Library collections and resources.

Complementing the NLM exhibitions and related educational initiatives are a number of HMD programs and external collaborations designed to raise scholarly, educational, and public awareness and use of the NLM rich historical collections which date from the 11th century to recent decades. Noteworthy initiatives during FY2014 included:

- The Division's blog, *Circulating Now*, which continued to offer rich, curated content about the NLM historical collections and the opportunity for readers to submit comments and exchange ideas about the posts. By the end of the year, *Circulating Now* had published over 300 posts and acquired an audience of thousands of direct subscribers and hundreds of thousands of followers.
- Joining the Flickr Commons, a global initiative among cultural heritage organizations, by offering a photo stream of public domain images from the HMD collections which are also available through the Images from the History of Medicine database.
- Leading the NLM hosting of the inaugural symposium of the National Digital Stewardship Alliance, entitled "Emerging Trends in Digital Stewardship." This free and public program included presentations by digital practitioners from across the Washington, D.C. area region on preserving social media and collaborative workspaces, open government and open data, and digital strategies for public and non-profit institutions. Releasing rare and historically-significant amateur footage of Franklin Delano Roosevelt speaking at the dedication ceremony for the NIH.
- Continuing its formal collection-lending program, by lending three items to The Grolier Club of New York City, for its exhibition *Extraordinary Women in Science and Medicine: Four Centuries of Achievement*. Adding two new profiles to the Profiles in Science Web site during FY2014, bringing the total to 38. The first new site, created in collaboration with the Osler Library of the History of Medicine at McGill University and the Alan Mason Chesney Medical Archives of the Johns Hopkins Medical Institutions, tells the story of the eminent Canadian physician Sir William Osler (1849-1919). The second new site is based on the career of Mike Gorman (1913-1989), a well-known journalist, author, publicist, and crusader for health policy reform. Gorman won a Lasker Award in 1948 for his newspaper exposés of state mental hospital conditions in Oklahoma. NLM is the repository for the Mike Gorman Papers, which range from 1946 to 1989.

Training and Recruitment of Health Sciences Librarians

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

In FY2014, the MEDLARS Management Section (MMS) and the NTC continued to provide a variety of training courses for librarians. Overall, 1,660 students were trained in 54 classes covering the subjects of PubMed for Trainers, PubMed for Librarians, "Teaching with Technology," and "A Librarian's Guide to NCBI." In addition, use of NLM distance education offerings continues to grow with 1,802,873 page views of the 65 online tutorials and Quick Tours available.

The NLM Associate Fellowship program had nine participants in FY2014: four 2nd year fellows at New York University, Washington University at St. Louis, Northwestern University, and the University of Arizona, and five 1st year fellows, who completed their year at NLM in August 2014. Second year placements were arranged for one of these fellows at the Mayo Clinic – Arizona, and two at Johns Hopkins University. One fellow took a permanent position at the University of Cincinnati, and one fellow took a position at the National Library of Medicine. Four new fellows began the 2014-2015 Associate Fellowship year at NLM in September. Efforts to recruit fellows from underrepresented groups have been successful in attracting diverse groups of fellows to the program, including Hispanic representation in two of the last six cohorts and African American representation in three of the last six cohorts.

NLM works collaboratively with the American Library Association and the Medical Library Association to provide funding for the ALA Spectrum Scholarship Program. NLM has supported the Spectrum Scholarship Program for 14 years, with a commitment to continue support for another six years. The ALA Spectrum Scholarship program recruits library science students from underrepresented populations and supports their academic and leadership development through tuition support and attendance at the ALA's annual meeting. NLM is also a partner library in an Institute of Museum and Library Services (IMLS)-funded initiative with the Association of Research Libraries (ARL), for the ARL Career

Enhancement Program (CEP). The ARL CEP supports library science students from underrepresented populations with paid internships at an ARL library and attendance at an annual Leadership Symposium. LO hosted two ARL CEP fellows in summer 2014. LO also supports the

NLM/AAHSL Leadership Development Program which provides leadership training, mentorship, and site visits to a mentor's institution for an annual cohort of five midcareer health sciences librarians. Recruitment efforts have emphasized and been successful in attracting minority candidates.

Table 1: Collections

<i>Physical</i>	<i>Total¹</i>	<i>FY2014</i>	<i>FY2013</i>	<i>FY2012</i>
Monographs ²				
Before 1500	598	1	0	0
1501-1600	6,058	2	3	0
1601-1700	10,347	1	3	0
1701-1800	272,741	3	3	4
1801-1870	256,788	71	27	37
1871-Present	886,034	13,560	11,292	13,287
Bound Serial Volumes ³	1,480,296	15,806	12,650	17,025
Microforms ⁴	606,126	42	8	10
Audiovisuals and Computer Software	95,356	2,395	1,701	1,310
Prints and Photographs	70,231	279	300	336
Manuscripts ⁵	23,096,447	4,807,740	914,025	1,716,225
Withdrawn Items	(132,761)	(549)	(350)	(294)
Total items	26,648,261	4,839,351	939,662	1,747,940
<i>Digital</i>	<i>Total¹</i>	<i>FY2014</i>	<i>FY2013</i>	<i>FY2012</i>
PubMed Central Articles	3,227,379	368,111	318,316	263,203
PubMed Central Titles ⁶	1,600	207	244	252
Bookshelf Titles ⁷	3,106	1,387	323	323
Digital Collections Repository ⁸				
Texts ⁹	12,201	2,642	2,580	5,075
Audiovisuals ¹⁰	162	55	37	41

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 website.

¹⁰ Referred to as "Films and Videos" on Digital Collections website.

Table 2: Collection Activities

<i>Acquisitions and Processing</i>	<i>FY2014</i>	<i>FY2013</i>	<i>FY2012</i>
Active Serial Subscriptions	17,439	18,343	19,184
Items Processed ¹¹			
Serial Pieces	94,738	99,891	101,294
Monographs (pre-1914) ¹²	1,218	336	715
Monographs (1914-)	19,367	16,530	17,012
Audiovisuals ¹³	241	683	1,335
Prints and Photographs	1,364	1,397	47,982
Total	134,367	137,180	168,338
<i>Archival Materials Acquired</i>			
Modern Manuscripts (in linear feet)	157	120	497
<i>Expenditures</i>			
Publications	\$11,571,597	\$11,033,522	\$10,207,330
Rare Books, Manuscripts, and other Historical Materials	\$299,841	\$299,948	\$299,584
Total¹⁴	\$11,871,438	\$11,333,470	\$10,506,914
<i>Preservation</i>			
Volumes Bound	14,516	14,903	15,000
Volumes Repaired Onsite ¹⁵	685	994	2,346
Audiovisuals Preserved	811	632	534
Historical Volumes Conserved	583	375	997
Pages Digitized ¹⁶	413,550	540,830	643,372

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	FY2014	FY2013	FY2012
General Collection Items ¹⁷	18,755	16,685	19,656
Historical Monographs (pre-1914)	4,431	4,342	4,238
Modern Manuscripts (in linear feet) ¹⁸	92	303	1,050
Prints and Photographs ¹⁹	1,326	2,083	2,281
Historical Audiovisuals	875	894	992
Indexing	FY2014	FY2013	FY2012
Citations Indexed for MEDLINE	765,850	734,052	760,903
Journals Indexed for MEDLINE	5,647	5,640	5,633

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

<i>Document Delivery</i>	<i>FY2014</i>	<i>FY2013</i>	<i>FY2012</i>
Interlibrary Loan Requests Received	188,912	218,268	234,662
Interlibrary Loan Requests Filled	155,423	179,941	194,255
General Reading Room Requests			
Received	68,281	76,215	85,352
General Reading Room Requests Filled	61,713	68,486	74,993
History of Medicine Reading Room			
Requests Filled	8,660	7,639	6,989
 <i>Customer Service Inquiries</i>	 <i>FY2014</i>	 <i>FY2013</i>	 <i>FY2012</i>
Offsite Inquiries ²⁰			
General	95,665	99,857	100,023
History of Medicine	5,776	8,108	10,857
Onsite Inquiries ²¹			
General	6,967	5,219	6,124
History of Medicine	9,750	12,183	14,763
 <i>Data Licensees</i>	 <i>FY2014</i>	 <i>FY2013</i>	 <i>FY2012</i>
MEDLINE	831	767	736*
UMLS	12,627	8,804	5,640
 <i>Tours and Visitors</i>	 <i>FY2014</i>	 <i>FY2013</i>	 <i>FY2012</i>
Exhibitions			
Visitors	2,700	2,662	2,769
Daily Tours			
Tours	83	111	122
Visitors	508	734	668
Special Tours			
Tours	75	88	63
Visitors	1,952	1,660	1,047

*This figure corrected from previous reports.

²⁰ Offsite Inquiries: Inquiries via telephone, fax, U.S. mail, and e-mail. Includes Bibliographic Services Division (BSD) interactions with data licensees.

²¹ Onsite Inquiries: In person

Table 5: Selected Web Resources

<i>Resource</i>	<i>FY2014</i>	<i>FY2013</i>	<i>FY2012</i>
<i>ClinicalTrials.gov</i>			
Number of Trials	176,622	157,013	133,291
Page Views ²²	1,414,637,991	1,145,603,153	1,474,668,417
Visitors ²³	11,264,641	11,058,828	11,201,854
<i>DailyMed</i>			
Number of Labels	66,527	55,190	39,232
Page Views	182,881,007	146,216,340	85,025,368
Visitors	18,722,781	17,541,406	8,812,640
<i>Genetics Home Reference</i>			
Summaries	2,376	2,125	1,907
Page Views	198,089,169	161,918,165	145,647,998
Visitors	14,064,110	9,719,450	6,382,520
<i>Household Products Database</i>			
Number of Products	14,000	12,000	11,000
Page Views	41,757,342	39,960,413	17,994,336
Visitors	970,643	820,718	786,346
<i>MEDLINE/PubMed</i>			
PubMed Citations	24,289,399	23,187,946	22,174,097
Page Views	8,094,826,103	7,573,459,637	6,113,756,836
Unique Visits ²⁴	741,750,623	716,895,000	482,175,373
Searches ²⁵	2,650,894,898	2,514,611,135	2,204,710,177
<i>MedlinePlus</i>			
Number of Topics (English/Spanish)	957/948	953/939	931/918
Page Views	1,023,100,000	845,000,000	745,000,000
Visitors	409,800,000	298,500,000	213,000,000
<i>NLM Main Web Site</i>			
Page Views	60,300,000	44,600,000	44,700,000
Visitors	14,500,000	9,100,000	8,900,000
<i>ToxTown</i>			
Page Views	5,443,181	5,029,713	6,757,011
Visitors	294,186	263,781	240,277

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) offers a wide-ranging collection of publicly accessible free online information resources and services in toxicology, environmental health, chemistry, HIV/AIDS, disaster management, minority health, and other specialized topics. The Division was created in 1967 with an initial focus on toxicology and chemistry and has subsequently expanded to include environmental health, drug information, and other specialized topics of significant interest to health professionals and the public.

In addition to a branch focused on developing the content of TOXNET and other core toxicology and chemical databases and a branch providing IT support, SIS includes an Outreach and Special Populations Branch, which actively seeks to improve awareness of and access to high quality health information by underserved and other targeted populations. SIS also manages NLM's Disaster Information Management Research Center (DIMRC). Its Office of Clinical Toxicology (OCT) works closely with other parts of the Division, NLM, NIH, and HHS, as well as with outside clinical toxicologists and other potential users. SIS has played a leadership role in the development of HIV/AIDS information resources since 1989. An outgrowth of our management of HIV/AIDS information resources has been the development and management of a large contract, Support for NLM and NIH Biomedical and Clinical Information Services, which supports work for Lister Hill National Center for Biomedical Informatics, the National Institute of Allergy and Infectious Disease (NIAID), the National Heart Lung and Blood Institute (NHLBI), and other components of NIH and HHS.

The SIS staff includes librarians, biomedical scientists, computer scientists, public health practitioners, evaluators and educators. The work of the Division involves developing databases, Web sites, tools for use by professionals and the public, training courses and instructional materials related to the SIS resources and programs. Staff uses all available tools to communicate with users and potential users of SIS and other NLM information resources and services, including Twitter feeds, Facebook pages, Instagram, Pinterest boards, memes, and infographics. We are assisted in these efforts by a robust group of interns mainly from local public health and information science programs who stay for a semester and gain college or graduate credit for their work.

In addition to the undergraduate and Masters level interns that SIS has been hosting, the Division also hosted two physicians from Uniformed Services University for practica.

SIS also uses traditional methods of scientific communication including posters and presentations at conferences and publication in journals. Presentations were as diverse as SIS resources and included "Snakes Found in Maryland, and Snake Bite Prevention and Treatment" at a Maryland State Park and lectures in the fall and spring in the toxicology courses offered to physicians and others by the Preventive Medicine & Biometrics Department at the Uniformed Services University of the Health Sciences (USU). Numerous staff members authored or co-authored papers including two that reported on aspects of the evaluation of the AIDS Community Information Outreach Program.

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 (toxicology and environmental health, 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. *Health Hotlines*, the very popular online database of health-related organizations operating toll-free telephone services, was revised and updated this year. It is now available for downloading from the updated Web site in printable pdf and eBook formats.

Toxicology and Environmental Health Resources

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established more than 40 years ago within the National Library of Medicine as a result of recommendations of the President's Science Advisory Committee. From its inception, TEHIP has strived to use the most up-to-date and relevant information together with the use of 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 sources of toxicological and environmental health information wherever these sources may reside as part of the library's role in collecting and organizing health and medical information. Identifying high quality resources is significant facet of the work done within SIS, and we continue to explore new ways to reach users with relevant authoritative information. We are exploring ways to better utilize visualization in our varied resources.

Screenshot of the TOXicology Data NETwork (TOXNET) websites's Homepage

Our *TOXNET* (TOXicology Data NETwork) <http://toxnet.nlm.nih.gov> databases are the core content in our TEHIP information resources. These databases are varied in what they cover, and how they are created and maintained. Some are legacy ones, evolving from their earlier forms as we included information on a wider range of toxic chemicals, such as adding Superfund cited compounds, focusing on consumer and worker exposures, taking on the possible safety issues with nanomaterials and considering household products safety. Others are databases originally created elsewhere and put up on the TOXNET system for greater visibility and to provide continued access when the project that created them was stopped. Enhancements to TOXNET continue, based on user feedback and upgrades or additions of data and capabilities. A new interface for TOXNET, as seen in the main search page shown above, was released in FY2014. User response to the new look and functionality has been positive, and we continue to improve the interface to help direct users quickly to needed information. Some of the most highly used databases that are currently part of the TOXNET system 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. The *LactMed* app for the iOS and Android continues to be very popular. Discussions are underway to increase the scope of the database to include drugs used during pregnancy.

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 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. Following the suggestions in the recent HSDB Needs Assessment project, we continued to work to better incorporate new resources important to today's toxicologists and other potential users. We held our first all-electronic review in September of 2014, and will evaluate how to most effectively incorporate this process in to the peer review of HSDB. Particularly noteworthy this year was the rapid development, review by the HSDB Scientific Review Panel, and deployment of information on the 4-Methylcyclohexanemethanol (MCHM) chemical release in

to the West Virginia Elk River in January 2014. The HSDB record for MCHM was immediately made available through CDC.

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 RePORTER system, allowing users to link directly back to the NIH database for full information on the projects. DART/ETIC (Developmental and Reproductive Toxicology/Environmental Teratology Information Center), a bibliographic database covering literature on reproductive and developmental toxicology. This database is no longer funded by the multi-agency group that created it, but it is still searchable as a distinct entity as well as a part of TOXLINE.

TRI (Toxics Release Inventory), a series of databases that describe the releases of toxic chemicals into the environment annually for the 1987-2011 reporting years. The data are compiled by EPA from required industry submissions.

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.

Household Products Database, which provides information on the potential health effects of chemicals contained in more than 14,000 common household products used inside and around the home. Collaboration continues with the Consumer Product Safety Commission (CPSC) to include some types of nanomaterials and additional categories of products in the Household Product Database. The latest update includes a new product category "commercial/institutional" which encompasses over 300 products that may be described by their manufacturers as professional grade, professional use, hospital grade and more.

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. Haz-Map added 1989 chemical and biological agents in FY2014, In addition twenty-three new agents causing occupational asthma, fifteen new hazardous job tasks linked to jobs and

industries, and several new agent categories were added. Haz-Map now includes 287 agent categories.

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 Alternative Methods (ICCVAM). ALTBIB provides numerous topic area searches of PubMed as well as easy access to key organizations, the latest news from some of those organizations, and to US and international compilations of validated methods and methods undergoing validation.

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

TOXMAP, a Geographic Information System (GIS) that uses maps of the United States to help users visually view data about chemicals released into the environment and easily connect to related environmental health information.

Drug Information Portal, providing current drug information for over 28,000 drugs with links to many credible additional online resources. We continue to improve the searching, display, and links of this resource.

Pillbox focuses on pill images as well as drug names and other physical appearance information. In 2014, Pillbox was selected to participate in the HHS Ignite program. During the program, Pillbox developed an open source data processing system to parse the Structured Product Labels from DailyMed and RxNorm, to create Pillbox's pill-focused data set. The staff and time resources required to perform the previous data process were cut by 95 percent. The collaboration between Pillbox team and the Department of Veterans Affairs (VA), Medication Image Library (MIL) resulted in the availability of high-resolution images and better quality control of the data. VA MIL images now comprise over 60 percent of the images available in Pillbox.

Dietary Supplements Label Database is created in collaboration with the NIH Office of Dietary Supplements (ODS). 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. SIS previously had licensed data to produce a database for consumers which is no longer available through NLM.

Comparative Toxicogenomics Database (CTD), developed at the Mount Desert Island Biological Laboratory and now located at North Carolina State University, is included in TOXNET searching. CTD receives funding from the National Institute of Environmental Health Sciences (NIEHS). This database focuses on molecular mechanisms by which environmental chemicals affect human disease,

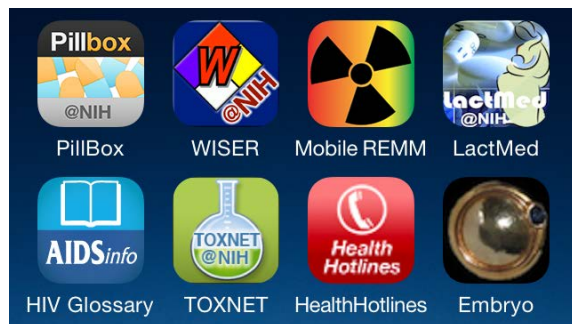
providing insight into complex chemical-gene and protein interaction networks.

LIVERTOX, a Web site that provides up-to-date, accurate, and easily accessed information on the diagnosis, cause, frequency, patterns, and management of liver injury attributable to prescription and nonprescription medications, herbals and dietary supplements, continues to be managed by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The site includes a case registry that will enable scientific analysis and better characterization of the clinical patterns of liver injury. LiverTox contains approximately 850 drug and herbal records.

Enviro-Health Links: SIS continued to update and enhance topic pages such as the ones on laboratory safety, and nanotechnology, and tobacco, smoking, and health. For example, the laboratory safety one has links to information for clinical, academic and school laboratories, including resources for handling chemical, biological and nanotechnology safely. Also included are links to regulations and policy, hazard analysis, Material Safety Data Sheets, and waste management, along with preformulated TOXNET and PubMed searches for likely terms and topics of interest.

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. SIS continues to enhance and develop mobile apps/mobile Web resources in the area of emergency response, health information, drug information, toxicology, and embryology. This year, Pillbox released the Pillbox App for the iOS platform. Mobile REMM released a new design with improved user interface design. In addition, *AIDSinfo* and Environmental Health Student Portal released the responsive designed Web sites to provide optimal user experiences across a wide range of devices such as mobile phones, tablets, and desktops.



Specialized Information Services Mobile Application and Web Resources Icons

Disaster Information

Just-in-time, just-what-I-need information continued to be one motto for the Disaster Information Management Center (DIMRC) in FY2014. DIMRC provided authoritative health information for disaster preparedness, response, and recovery and to assist with information needs and dissemination for specific incidents including the chemical spill in the Elk River, West Virginia and the Ebola outbreak. The most serious outbreak of Ebola to-date began in West Africa the spring of 2014 and quickly grew into a public health emergency as defined by the World Health Organization. In September, the first case in the United States was diagnosed. Fear of Ebola spread rapidly and the need for accurate, timely information for health professionals and the public was critical. DIMRC staff quickly pulled together information and created a Web page with links to key sources of guidance including CDC, WHO, APHA, and many more. In addition, records and links to all guidance documents, training resources, Web pages, reports and Congressional hearings were added to the Disaster Lit: The Resource Guide for Disaster Medicine and Public Health database, for easy identification and retrieval by public health staff, health professionals, and state, local, and government agencies. The height of the outbreak spanned multiple fiscal years and approximately 100 resources on Ebola were added to Disaster Lit by the end of FY2014, over 500 resources on Ebola were added to Disaster Lit by the end of calendar year 2014. DIMRC staff used its Disaster Outreach listserv, related listservs, twitter, email, Webinars, and other methods to disseminate authoritative information on the outbreak to thousands of people.

Disaster Information Outreach

The Disaster Information Specialist (DIS) Program is a collaborative effort to explore and promote the role of information specialists/librarians in the providing access to disaster-related information resources to the disaster medicine and public health professionals and the communities they serve. In FY2014, DIMRC hosted nine Webinars that serve as continuing education programs for librarians and members of the disaster workforce. These ongoing Webinars 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 DIMRC staff with an update on DIMRC resources and presentations from several subject matter experts on real-time mapping of disease outbreaks, special needs and information resources about children in disasters, accessing information about chemical hazards (in response to the chemical spill in WV), public health preparedness training resources, and use of social media and mHealth tools by local health departments. In addition, presentations by the recipients of NLM funding for disaster outreach projects were given during the summer. The number of attendees

increased from an average of 60 callers per Webinar to 100 per Webinar (we do not know if multiple people participated from one location). The Webinars are recorded and transcribed for viewing at any time.

DIMRC continued hosting the Disaster Information Specialist courses and offering the DIS basic and advanced certificates through the Medical Library Association (MLA). A new course on Information Resources for Chemical, Biological, Radiological, and Nuclear Incidents was developed and taught at the annual MLA meeting in May. Materials were transferred from the Medical Library Association's Web site to the DIMRC Web site last year and this year SIS began transitioning the courses to a new Learning Management System.

The Disaster Health Information Outreach program continued to grow and at the end of FY2014, NLM awarded purchase orders to four new organizations. The projects involve an organization working in disasters (public health department, emergency management department, fire/rescue service, academia, etc.) to partner with a library to improve access and knowledge of NLM and other disaster health resources and to identify roles for librarians in disaster management. The new projects focus on training emergency management professionals to use disaster information resources, including apps and mobile Web sites in Idaho, and two cities in Indiana, and the development of an app at the University of Washington. Three projects were completed during FY2014 from awards made the previous year. These projects included training staff of the 211 center and the regional office of the Pan American Health Organization in El Paso, Texas, teaching health professional and librarians in American Samoa to use NLM DIMRC resources, and information management training for emergency management staff in Fort Worth, Texas.

Promoting awareness of and access to disaster health resources continued to be a priority in FY2014. The Disaster Outreach listserv membership grew from over 1,100 subscribers at the end of FY2013 to nearly 1,300 at the end of FY2014. New members include both librarians and the disaster medicine and public health workforce. DIMRC continued to use twitter extensively to reach many new people and to assist other agencies, partners, and collaborators in sharing and disseminating information more widely. DIMRC staff kept up with current events and activities and helped communicate timely information to a wide audience, especially promoting new guidance documents for management of Ebola patients and highlighting new resources added to Disaster Lit. The @NLM_DIMRC twitter account nearly doubled its number of followers to almost 3,000 at the end of FY2014. Twitter allowed us to quickly and easily share information about DIMRC and other NLM resources, especially after the chemical spill in West Virginia and the Ebola outbreak, and to collaborate with other agencies, organizations, and people involved in disaster medicine and public health such as the Assistant Secretary for Preparedness and Response (@PHEgov), the National Association of County

and City Health Officers (@NACCHOalerts), and the Director of the Disaster DistressLine. Over 2,500 tweets were sent in FY2014. The tweetbank to store "evergreen" tweets greatly aided our ability to tweet frequently and maintain an active presence in the disaster management twitter arena. According to Bhattachyra et al (2014) (<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0112235>), @NLM_DIMRC had the highest number of favorites of any HHS twitter handle.

Presentations at meetings and conferences continued to be useful in teaching users about DIMRC and other NLM resources. For example, this year DIMRC was well represented at the 2014 Preparedness Summit through teaching a class on How to Navigate the Information Maze to Access Evidence-based Public Health Preparedness Information, giving a presentation on REMM in a session on radiation preparedness, participating on panels on social media for situational awareness and on continuing education for emergency managers, presenting a poster on roles for librarians in disaster information management, and participating in a podcast on the use of social media in disasters. SIS staff also presented disaster health information at other meetings, including the Medical Library Association, the Medical Reserve Corps of Loudoun County, VA monthly meetings, the Texas Library Association annual meeting and numerous Webinars.

The Latin American Disaster and Health Information Network (LANDHI or RELACIGER in Spanish), continued working together and several centers completed health IT projects with funding from DIMRC. Through these projects, LANDHI developed a system to monitor the availability of each member's Web site over time, an enhanced mobile application for Android and iOS, two additional national networks that collaborate with the local LANDHI center, a digital collection of disaster management videos, a collection of disaster-related legislative resources, and a network bulletin that publishes and disseminates pertinent articles written by the LANDHI community. LANDHI members also conducted disaster information workshops in Nicaragua and El Salvador and developed an enhanced Web site for the network. Members meet via Webinar several times a year to discuss activities and plans. LANDHI includes 16 disaster and health information centers in 11 countries.

At the end of FY2014, staff traveled to Haiti to meet with the country representative of the Pan American Health Organization and with staff from several government agencies to discuss the transfer of the Haiti Disaster Information Center Web site and database from Costa Rica to Haiti. SIS funded the creation and maintenance of the Web site which was being maintained in Costa Rica until Haiti was prepared to take responsibility for it.

Disaster Health Information Resources and Tools

The name of the grey literature database was modified to Disaster Lit: The Resource Guide for Disaster Medicine

and Public Health based on a usability study. Disaster Lit is shorter, easier to remember, and more clearly describes the database. Nearly 2,000 resources were added this year for a total of 7,000 records. As expected, in September a number of government agencies and non-governmental organizations began publishing a large number of resources on Ebola. All of these resources were added to Disaster Lit and can be easily retrieved with one click and filtered by organization, type of resource, etc.

DIMRC was approached by several agencies to assist with the management of existing collections or the development of new collections for Disaster Lit. DIMRC worked with the Health Resources and Services Administration (HRSA) and the Emergency Medical Services for Children (EMSC) to transfer the PedPrepared database from the EMSC site to Disaster Lit as the organizations were unable to sustain the database and did not want to lose this information. In addition to merging PedPrepared into Disaster Lit, a new Web page on Health Resources on Children in Disasters and Emergencies was developed in conjunction with HRSA and EMSC. The Web page contains searches of Disaster Lit and PubMed on a variety of topics including behavioral health, chemical biological radiological and nuclear enhanced (CBRNE) weapons, pandemics, active shooter and the special needs of children.

This year, NIEHS and NLM DIMRC collaborated on a pilot NIH Disaster Research Response project with funds from the NIH Director's Discretionary Fund. This project is investigating mechanisms to conduct research following a disaster in a timely fashion. DIMRC initially developed a Web page about the project that will migrate to its own Web site in FY2015. In addition, DIMRC worked with NIEHS to identify disaster research tools (surveys, questionnaires, etc.) that have been used following disasters. These tools were added to the Disaster Lit database and will be available and searchable as part of the new Web site.

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). The new interactive tool has been used in the Secretary's Operation Center (SOC) for several incidents including the chemical spill in West Virginia and the Ebola outbreak. 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 and ASPR are working to integrate incident-specific annexes (e.g., hurricane, earthquake, chemical emergency) into the tool.

NLM collaborated with the Food and Drug Administration (FDA) on the development of a mobile app to collect data on the adverse effects of drugs or devices approved via Emergency Use Authorization in declared public health emergencies. The Lister Hill National Center for Biomedical Communications provided assistance and guidance on the app development, and SIS conducted several usability tests of the app at various stages of

development as well as the dashboard to collect and analyze the data.

HazMat/CBRN Tools

WISER and CHEMM: Significant improvements were made to several WISER platforms including webWISER, WISER for Windows and WISER for the Android. Our users recommended integrating WISER and CHEMM information, and this was accomplished last year for the iOS and Android apps. In FY2014, this was done for WISER for Windows and webWISER. In addition, these platforms were updated with the most current version of the Department of Transportation's Emergency Response Guidebook (ERG). Both the popular "Help Identify" and "Protective Distance Mapping" features were added to the Android platform, making the Windows, iOS, and Android platforms equivalent. In addition, the iOS platform was updated to provide full support for iOS7 and above with a new look and feel to match the latest iOS releases.

DIMRC continued to work with the Department of Transportation's Pipeline and Hazardous Materials Safety Administration (PHMSA) to maintain the ERG 2012 app for iOS and Android. DOT requested NLM's assistance due to our expertise with app development and maintenance.

Outreach

SIS continued a robust set of projects and programs intended to promote use of health information resources in a diverse array of community and professional settings. 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. The OSPB Web site (<http://sis.nlm.nih.gov/outreach.html>) provides access to many specific population resources and programs. In addition, we use varied channels including Twitter, Facebook, and listservs to communicate with our intended audiences.

Environmental Health Information Partnership (EnHIP)

EnHIP, NLM's long standing partnership with minority serving institutions, endeavors to enhance the capacity of the institutions to reduce health disparities through the access, use and delivery of environmental health information on their campuses. EnHIP, now in its 23rd year, has 22 partnering institutions including 14 HBCUs, three tribal colleges, three Hispanic-serving institutions, one Alaskan university and one community college. The FY2014 meeting, entitled NLM in Action included a keynote address, *The Human Microbiome: What It Is and How It Affects Your Health* by Lita Proctor, PhD, Project Coordinator, NIH Human Microbiome Project, National Human Genome Research Institute, NIH. John Koerner,

MPH, CIH, Chief, CBRNE Branch, Tactical Program Division, Office of Emergency Management, HHS, provided *Science Response-How NLM Supports Preparedness and Emergency Operations*. NLM staff provided updates on TOXNET data and its new Interface, NIH Big Data to Knowledge Initiative, NIH and NLM grant programs, the Office of Clinical Toxicology at SIS, and current status and informatics research challenge around pill images. Dr. Ann Barbre, Xavier University, chaired the meeting. Nine outreach projects were funded with EnHIP partner institutions in FY2014.

Community Outreach

Community Health Maps blog shares information about free/low cost and easy-to-use Geographic Information Systems (GIS) tools which can help community-based and other small organizations collect and visualize data such as health statistics, demographics, and community resources and events (communityhealthmaps.nlm.nih.gov). This information can be used to support community health planning and decision-making and to facilitate a better understanding of community conditions. The interactive nature of blogging helps Community Health Maps share information about hardware platforms and software applications available to communities as they consider how, or if, they might use GIS.

SIS continued its work with Black Greek Letter Organizations. These organizations are involved in local outreach efforts to combat the health disparities that adversely affect minorities. Staff worked with members of Phi Beta Sigma Fraternity to promote health literacy and health awareness.

NLM4Caregivers program addresses the health information needs of caregivers across the life span. As part of the outreach efforts, staff completed four videos on MedlinePlus, pilot tested five class modules as part of an educational endeavor, and continued outreach efforts with the Ronald McDonald House in Dallas, Texas.

Petersburg Public Library in Virginia was awarded funding for the Healthy Living and Learning Center. The Library staff will develop programs for community physician and healthcare providers emphasizing women as the main information gatherers and health decisions influencers of the family.

Outreach to Health Professionals

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

Association of Hispanic Nurses conference, 7th Biennial Asian American, Native Hawaiian, and Pacific Islander Health Conference, National Indian Health Board, USET – United South and Eastern Tribes conference and the Sixth Annual National Conference on Health Disparities.

The EquityRx Health Equity Scholars Program aims to provide graduate students with opportunities to learn about health disparities and current efforts to eradicate them through research, education and policy analysis. Ten Equity Rx Health Equity Scholars participated in the program and received NLM database training before beginning their Equity Reports.

Outreach to Librarians

NLM continues to support the training and education of minority librarians through funding of a graduate assistantship for a student in the Knowledge River Program at the School of Library and Information Science, University of Arizona. The student works in the Health Sciences Library and is encouraged to pursue a career as a health sciences librarian.

NLM provided support for a health and wellness information project with the Black Caucus of the American Library Association. The Black Caucus of the American Library Association (BCALA) serves as an advocate for the development, promotion, and improvement of library services and resources to the nation's African American community and provides leadership for the recruitment and professional development of African American librarians. This project will address the health concern of obesity.

Special Population Web Sites

Arctic Health (<http://arctichealth.nlm.nih.gov>) the information portal that addresses the issues affecting the health of our planet's northern-most inhabitants was enhanced this year. A new section called "Multimedia" was added to collect and highlight the unique images, videos, and movies contained in Arctic Health. Also created was the "Publications and Research" section which will be enhanced with datasets and researcher profiles to help build partnerships with universities and research centers in the Arctic. Arctic Health staff from the University of Alaska, Anchorage published the paper "Sources of Alaska Native Health Data and Statistics" in the *Journal of the Canadian Health Libraries Association*. The article highlights Arctic Health as a source for statistics and information for and about Alaska Natives and other indigenous populations in the Arctic region.

Asian American Health is the information portal that addresses the health and well-being of Asian Americans in the United States (<http://asianamericanhealth.nlm.nih.gov/>). A poster entitled "Recommendations to Address Information Gaps on Asian American Health: Results from Project AXIS (Asian American Exchange

and Information Source)” outlined the results of the needs assessment and highlighted the informational needs of gatekeepers serving Asian American communities. It was presented at the American Public Health Association meeting with our partner New York University, Center for Study of Asian American Health. Based on the needs assessment, the “Materials in Asian Languages” section of the Web portal was removed to eliminate duplication of resources available in other NLM resources.

Women’s Health Resources portal links to significant research, initiatives, and topics in women’s health and sex and gender differences (<http://womenshealthresources.nlm.nih.gov/>). This year, the @WomensHealthNIH Twitter account (<https://twitter.com/WomensHealthNIH>) was retired based on results of a user needs assessment which found that our target audience of researchers does not use Twitter or Facebook for sex and gender research updates. The Women’s Health Facebook account (<http://www.facebook.com/womenshealthresources/>) was transferred to ORWH for future updates and information. New content on the WHR included presentations from the NIH Intimate Partner Violence Symposium. SIS continued its work with ORWH to increase the capacity for sex differences research by universities and increase knowledge and use of relevant research among students, faculty, and health professionals. Arizona Health Sciences Library, University of Arizona; Earl S. Richardson Library, Morgan State University; Lamar Soutter Library, University of Massachusetts Medical School; Lister Hill Library of the Health Sciences, University of Alabama; Medical University of South Carolina Library; Spencer S. Eccles Health Sciences Library, University of Utah; and Health Science Center Libraries, University of Florida received funding for related projects.

The Refugee Health Information Network (<http://www.rhin.org>) database of health education materials in multiple languages has been renamed HealthReach (<http://healthreach.nlm.nih.gov>). With this change, HealthReach is now an information resource to assist health care professionals and others working with communities with limited English proficiency. HealthReach will continue its collaboration with the Association of Refugee Health Coordinators, NLM and the Center for Public Service Communication (CPSC) to add and review resources.



Image Associated with the HealthReach Web site

Social Media

SIS continued its extensive use of Facebook and twitter for promotion and outreach. The NLM_HIVplus50 Twitter account started in January 2013 has almost tripled its followers and received recognition by AIDS.gov as a resource for information on HIV/AIDS and aging issues.

The social media team manages several Twitter feeds including:

- @NLM_SIS
- @NLM_OSP
- @NLM_DIMRC
- @NLM4Caregivers
- @AIDSInfo
- @HealthReach
- @NLM_HIVplus 50
- SIS K-12 Initiatives and Resources

K-12 Programs and Resources

NLM Distance Learning program is a partnership of NLM, 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 multi-way video. The 2013-2014 school years theme was Forensic Science. Presentations were provided by NLM and NIH personnel and others. This year a pilot program began with the staff at the University of Puerto Rico and Research Centers in Minority Institutions Program (RCMI) students. Site visits were made to Alaska and Hawaii for program evaluation.



Students from Hawaii, Alaska and Los Angeles communicating as part of NLM's Distance Learning Program

SIS coordinates the NLM Adopt-a-School Program, a partnership with the NLM Office of the Director, Woodrow Wilson Senior High School in Washington, DC and Charles H. Flowers High School in Maryland. 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. NLM hosts included Specialized Information Services, Library Operations, Office of the Director, and Lister Hill Center. The students made many contributions to NLM programs and activities.

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 and healthy behaviors, and develops leadership skills through student-community engagement. In March SIS hosted students from St. John's High School, John's Island, South Carolina who participated in the program. The students presented an overview of their school and community, presented on their outreach project for the year: Alcohol and Drugs: How they affect our community and engaged with the audience about future plans.

In addition, SIS initiated an in-depth qualitative evaluation of THLP. The evaluation focused on two cohorts of THLP scholars: the graduating classes of 2012 and 2013. Eleven of the fourteen former cohort scholars participated in semi-structured interviews. The evaluation documented the program's positive impact on the participants' health information competency, leadership skills, academic orientation, and interest in health careers. A paper discussing the findings has been accepted for publication in the March 2015 issue of the *Journal of Medical Library Association*.

Among several new health information literacy projects initiated this year, NLM, the Center for Public Service Communication and Talbot County Maryland High School are developing a teen leadership and mentoring project based on the South Carolina Teen Leadership Program model.

ToxTown, a resource about environmental health concerns where you live, work, and play, is undergoing an upgrade to allow for use on mobile devices. In addition, each scene will be refreshed to have a more photorealistic look. The *ToxTown* curriculum unit, *The Great Debate*, was introduced in a workshop at an NIEHS environmental literacy conference in North Carolina. Building upon the success of the *ToxTown* Chemical of the Week social media campaign, the K-12 team launched #toxloc on Twitter. The weekly tweet focused on an environmental concern in the various locations found in the *ToxTown* neighborhoods. Staff continued to work with the Estonian government to help them construct their *ToxTown* Estonia by sending files to be translated into Estonian.

Environmental Health Student Portal is a Web site for middle school students and teachers, which was enhanced in FY2014 in response to feedback from a usability study and focus groups. The responsively redesigned Portal now includes animated short videos with entertaining characters to convey environmental health ideas and concepts. Animations introduce middle school students to physical and chemical properties of mercury, pesticides, ozone, lead, and particulate matter, describe their use in industry and agriculture, and explained their impact on the environment and human health. The *Mercury and Our Health* video was accepted to be shown at the American Public Health Association annual conference.

In addition, work began on an educational game called *TOXInvaders*, based on the K-12 resources. In the game, users have to eliminate rapidly falling toxic chemicals, while collecting "good chemicals" (e.g., oxygen), through a succession of levels of gradually increasing difficulty. Proceeding to the next level requires passing a quiz drawing on SIS K-12 resources for information. In FY2014, SIS developed the storyline and general specifications of the game, built the prototype, and conducted a usability test with a group of middle school students.

GeneEd is a genetics information resource developed through collaboration within NLM (SIS, NCBI, LHCNCB) and the National Human Genome Research Institute that is intended for use by high school students and teachers. Substantially expanded during FY2014, *GeneED* now includes more than 650 different items including over 150 lesson plans and nearly 100 labs. *GeneEd* includes 56 genetics/genomics topics such as Cell Biology, Careers in Genetics, DNA, Genes, Chromosomes, Heredity/Inheritance patterns, Epigenetics/Inheritance and the

Environment, Genetic Conditions, DNA Forensics, and Top Issues in Genetics.

Internship Program

SIS has a strong internship program involving students from universities at both undergraduate and graduate levels as well as high school students and teachers through NLM’s Adopt-a-School Program. In FY14 SIS hosted 13 interns. Six of these worked primarily with the K-12 program. Jointly, they produced six educational animations for the NLM’s Environmental Health Student Portal. The interns also contributed to other SIS activities. One of the interns assisted with an evaluation of the NLM Teen Health Leadership Program and with a usability study of the HealthReach Web site (formerly RHIN). A high school student and an engineering student developed a prototype of a mobile game on genetics, and the high school teachers and the engineering student developed and tested an engineering activity to supplement water treatment information in the Environmental Health Student Portal. They also co-authored a manuscript about the engineering activity, currently in press in the Science Scope magazine of the National Science Teacher Association.

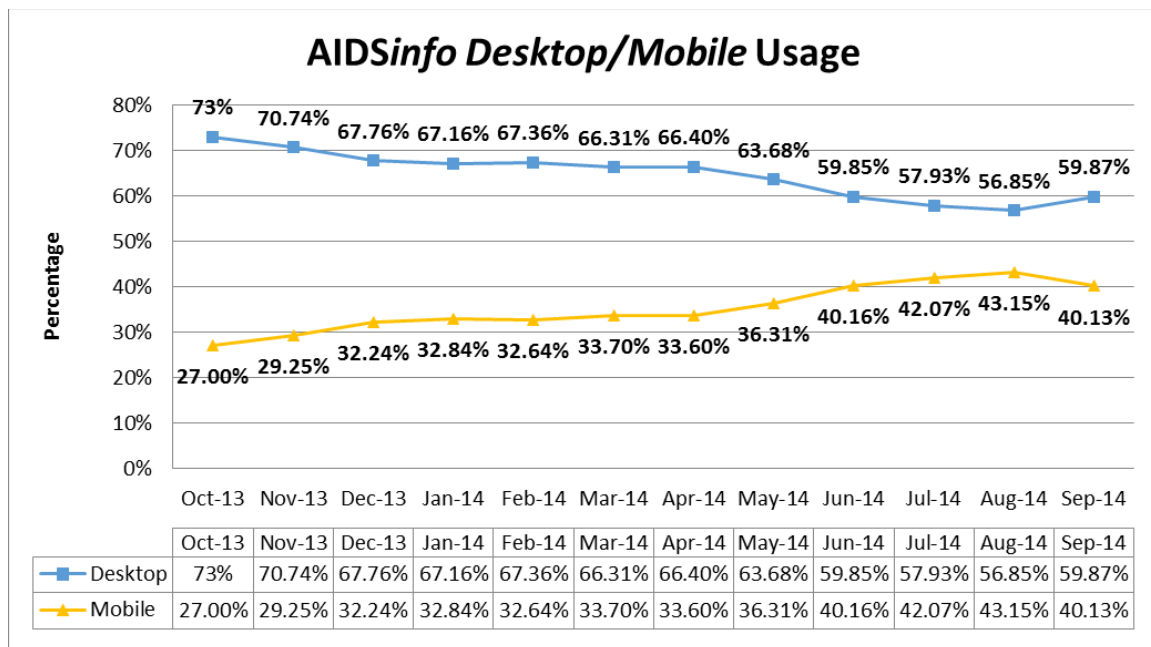
Seven students worked primarily with the social media team and the Outreach Branch. Jointly, these interns produced more than 365 Tweets, 200 Pinterest posts, and 150 Facebook posts about NLM resources and quality health information aimed at supporting family caregivers and family caregiver advocates. In addition, they

developed a process to maximize local outreach efforts, locating and attending five local outreach events. Their outreach efforts also included attending multiple Twitter Chats in which they promoted NLM resources online, in real time. Not only did the interns locate and disseminate quality health information to support our outreach efforts, they also educated SIS staff about emerging technologies.

HIV/AIDS

AIDSinfo/infoSIDA

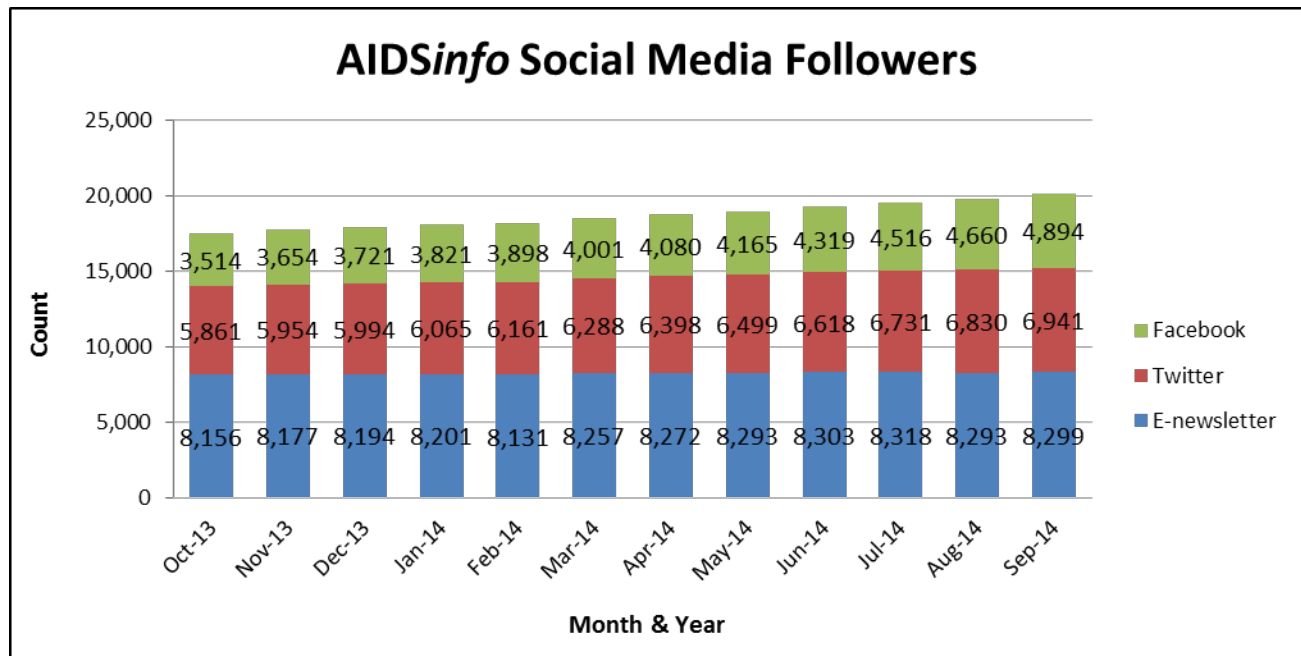
AIDSinfo is the HHS Web site for HIV/AIDS treatment, clinical trials, and prevention research information. In May 2014, the *AIDSinfo* and *infoSIDA* Web sites were launched in a responsively designed format. The Web sites are now automatically optimized for display across all devices, so visitors can access all of the Web site content on any device they are using (i.e., desktop computers, tablets, mobile phones – see image). This was a high priority update as increasing numbers of users are accessing AIDSinfo using mobile devices. The American Customer Satisfaction Index (ACSI) survey continues to show consistently high scores with an 81 at the end of the fiscal year. The *AIDSinfo* ACSI survey is now customized based on the device type the visitor is using to access the Web site and displays the corresponding survey questions (mobile vs. desktop). The AIDSinfo staff also supported the panels updating four of the treatment guidelines and updated the content on the Web site as well.



Desktop and Mobile Usage for AIDSinfo



AIDSinfo Responsive Design Web site



AIDSinfo Social Media Followers

HIV/AIDS Community Information Outreach Program (ACIOP)

NLM continued its HIV/AIDS related outreach to community-based organizations, patient advocacy groups, faith-based organizations, departments of health and libraries. Begun in 1994, The ACIOP 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. Eight projects were funded in FY2014. Staff introduced the ACIOP Blog (<http://aciopblog.wordpress.com/>) for sharing information about ACIOP outreach projects, resources, and services as well as to exchange ideas and build connections among the awardees. All Webinars, presentation slides, training manuals, sample evaluation plans, and documents relevant to the ACIOP are posted on this site. The FY2014 recipients served as the first cohort to participate in an enhanced evaluation process through data submission in reports and evaluations.

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). Through its biomedical informatics research, LHNCBC develops advanced health information resources and software tools that are widely used in biomedical research and by health IT professionals, health care providers, and consumers. 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, NIH institutes, and organizations within the Department of Health and Human Services (HHS), as well as with 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 semiannually to review LHNCBC's research projects and priorities. News and information about LHNCBC research activities are available at <http://lhncbc.nlm.nih.gov/>.

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 clinical care and medical training;
- Develop multimedia image-text databases that accentuate database organization, indexing, and retrieval; and
- Develop content-based image-retrieval (CBIR) techniques for automated indexing of medical images by image features.

Screening Chest X-rays for Tuberculosis and other Diseases in Rural Africa

In FY2014, we continued our collaborative project with AMPATH (Academic Model Providing Access to Healthcare), an organization supported by the US Agency for International Development (USAID) that runs the *largest AIDS treatment program in the world in Kenya*. 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.

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 chest X-rays (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. We explored many options for acquiring CXR training sets. We acquired 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. We curated the data sets from China and *Montgomery County and made them available for access by researchers worldwide*.

We provided AMPATH with lightweight digital X-ray units that are easy to transport in rural areas. The AMPATH staff is taking CXRs of people to screen them for the presence of lung disease. This project was selected for HHS Ignite, an initiative of the Department of Health and Human Services (HHS) Innovation Council, and received a cash award that was used to acquire field-deployable equipment. The award identifies projects that have the potential of making radical and positive improvements to the state of the art.

In the past year, one of our X-ray units was mounted onto a truck at the MOI University Hospital in the town of Eldoret in western Kenya. The images from the X-ray units are acquired in a standardized DICOM (Digital Imaging Communications) radiological-image

format. Through advances in technologies for Web-based Access to DICOM Objects (WADO) and the implementation of long-range Wi-Fi 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 and housed at the AMPATH building on the hospital grounds in Eldoret.



(Left) Mobile X-ray unit with generator and on-ramp deployed. (Right) The first patient of the day is prepped for a chest X-ray in Turbo, Kenya in September 2014.

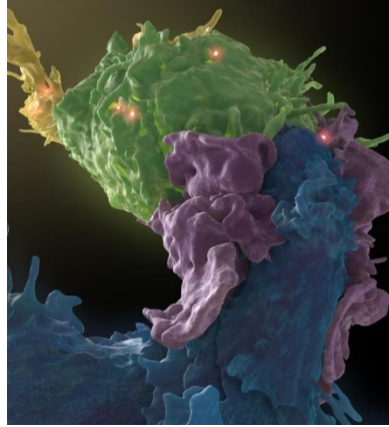
During FY2014, we were consulted by the National Institute of Allergy and Infectious Diseases (NIAID) about adapting our algorithm to detect pediatric tuberculosis. As a first step, we adapted the lung-segmentation algorithm to include pediatric lung shapes and found that it is reasonably accurate. We need a much larger training and test set to provide meaningful measures of performance quality. In 2014, we speeded up our lung-segmentation method, which can now segment lungs in tens of seconds rather than 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 Support Vector Machine (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 about 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.

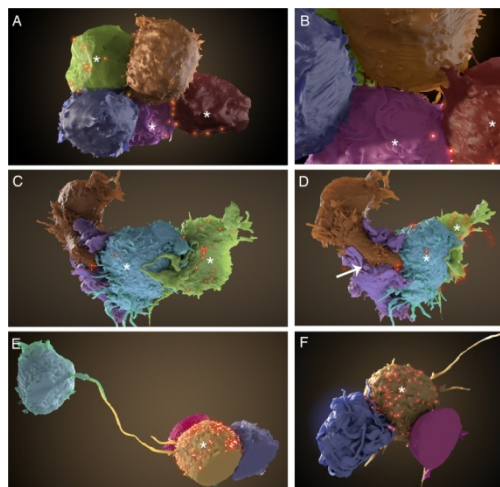
3D Informatics for High-Resolution Microscopy

We continue to address problems encountered in the world of three-dimensional and higher-dimensional, time-varying imaging through our 3D Informatics (3DI) and Molecular Visualization programs.

Throughout FY2014, we continued our collaboration with NCI's Laboratory for Cell Biology and with teams within LHCBC to visualize and analyze complex 3D volume data generated through dual-beam (ion-abrasion electron microscopy) and cryo-electron tomography. This work applies high-performance computing and audiovisual didactic development to data from 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. In collaboration with NCI, we developed a 3D illustration of HIV-1 transmission, which was featured on the cover of the *Journal of Virology*.

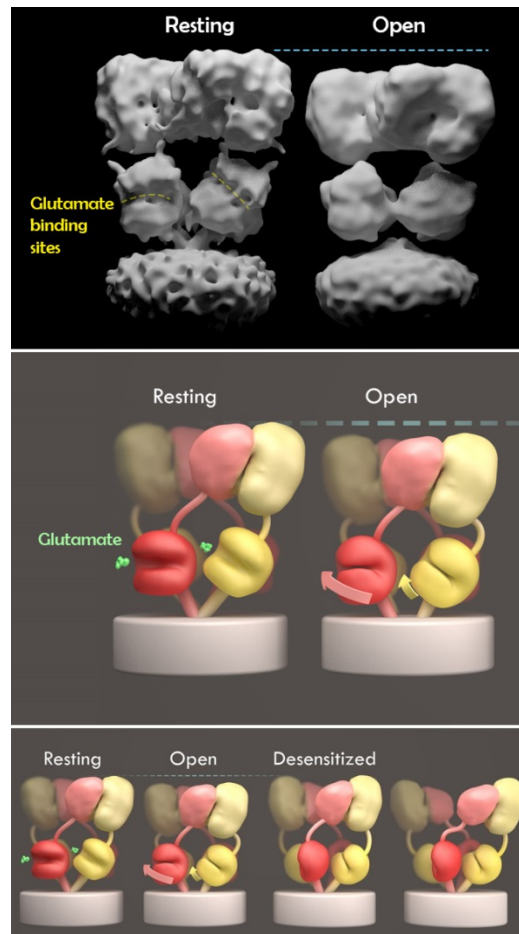


Uninfected blue and purple T-cells push through their environment to reach the green and yellow HIV infected T-cells in an image featured on the cover of the *Journal of Virology*. The work describes morphological differences in the virological synapses of a T-cell cell line (Jurkat cells) and primary human CD4 T-cells from peripheral blood. (*J Virol.* 2014 Sep; 88 (18):10327-39.)

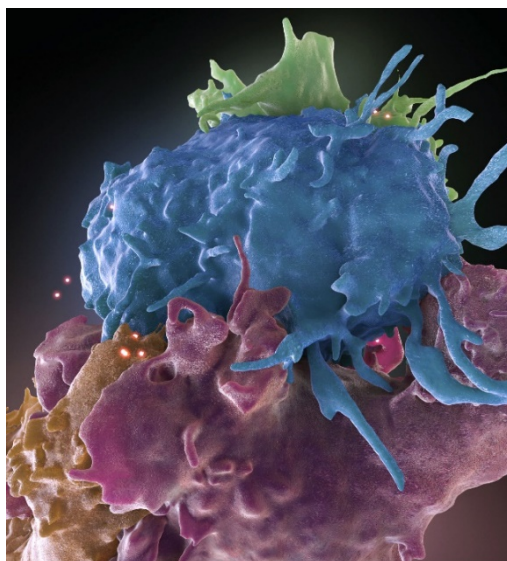


Segmentation of volume cell data shows morphological (shape) differences between common lab grown T-cells (top row) and T-cells from infected host. These differences could have an impact on vaccine research. Red spheres represent surface-bound HIV. (*J Virol.* 2014 Sep; 88 (18):10327-39.)

NLM worked with NCI to create a 3D animation of the HIV infectious spike. The 3D animation revealed a formerly undescribed twisting motion that affects binding to target membranes. Another collaboration with NCI clarified the structural mechanisms of glutamate receptor activation and desensitization in a *Nature* paper. We also used 3D-imaging software to translate data from cutting-edge NCI microscopy into illuminating visuals that led to new discoveries about influenza, including a publication about the feasibility of a universal influenza vaccine.



Stylized 3D model we developed in collaboration with NCI shows the discovery of relatively large molecular movements in the desensitized state of the glutamate receptor that are much harder to interpret directly from the data.



LHNCBC created this image, which features HIV-infected T-cells. (J Virol. 2014 Sep; 88(18):10327-39.) The New York Times requested and published it to illustrate a virus in their article entitled “Ebola and the Vast Viral Universe,” published October 28th, 2014.

We continued our commitment to processing data collected through transmission electron tomography. We successfully published the results of software-development research that uses graphics processing units (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. These 3D data are collected with resolutions of 3 nanometers across a field of view that can encompass entire 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.

Content-Based Image Retrieval (CBIR)

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. A goal of this research is 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 allows users to access 1.77 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 our 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*. The *Open-i* project is described in more detail below.

Authors place arrows on figures to highlight important regions. To help people find the images they need more quickly, we extract specific regions of interest (ROIs) within images. By analyzing the image layout, we’ve improved the performance of algorithms we developed for detecting and extracting the arrows. We’re developing additional algorithms to retrieve images that are more relevant to a query, whether the query is

submitted as text, a photo, or a combination. 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.

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."

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 ASCUS-LSIL Triage Study (ALTS) 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 and HPV-infection 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. In FY2014, we worked with Lehigh University to further develop this capability in the Multimodal Cervix Classification System (MCCS), which has the goal of accepting as input a uterine cervix image, with additional optional patient information such as HPV status, and producing as output a classification of the image into a low-risk or high-risk category.

In FY2014, in collaboration with NIAID, we developed an early prototype for automatically counting infected cells in blood films to assist malaria diagnostics, including cell detection and classification, and we finished implementing an image-processing method for measuring closing rates of cell-migration assays to support basic research of critical pathways for tissue repair. This method relies on many other methods used in the CBIR research area, including identifying cell boundaries and determining which cells are infected. Our malaria-imaging research received the Best Abstract Award at the Third Annual Seminar on Molecular Imaging of Infectious Diseases.

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 healthcare outcomes for people with cervical cancer, we plan to use these segmented epithelial regions to train doctors to detect various stages of precancer.

Imaging Tools for Biomedical Research and Education

In collaboration with cancer researchers and clinicians who conduct screening and diagnostic tests for cervical cancer, LHCBC conducts research and develops innovative biomedical imaging tools.

The American Society for Colposcopy and Cervical Pathology (ASCCP) continued to use one of our image-based systems, the Teaching Tool, to assess the knowledge and skills of colposcopy professionals. More than 130 resident programs in Obstetrics/Gynecology and Family Practice at more than 100 universities and other institutions, such as the Mayo Clinic, Baylor College of Medicine, Duke University, and the Tripler Army Medical Center, have used the tool. Since we first released the Teaching Tool in May 2010, these programs have administered more than 2,000 individual online exams (the Residents' Assessment of Competency in Colposcopy Exam (RACCE)); in addition, the ASCCP has used the Teaching Tool to administer the more-advanced Colposcopy Mentorship Program (CMP) exam to more than 370 practicing colposcopists.

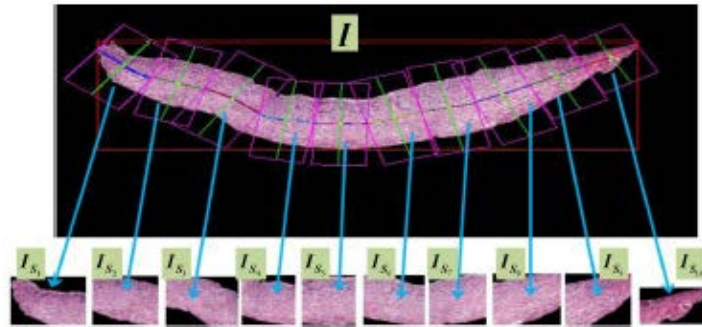
In 2014, we worked with the ASCCP to build more content for these exams by supporting the collection of expert colposcopist interpretations and biopsy regions for a set of 415 uterine cervix images. We plan to add the subset of these images on which the experts reached consensus opinion on quality, visual diagnosis, and biopsy region to the pool of image-based questions used by the RACCE and CMP exams. We also worked with the ASCCP to take steps toward the creation of an additional exam, the Colposcopy Recognition Award (CRA) exam, which is to test the highest level of expertise recognized by the ASCCP. We put a first, trial version of the CRA online in 2014 for ASCCP evaluation.

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 Health Sciences Center. 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, deriving image features based on this segmentation, and using the features in Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA)

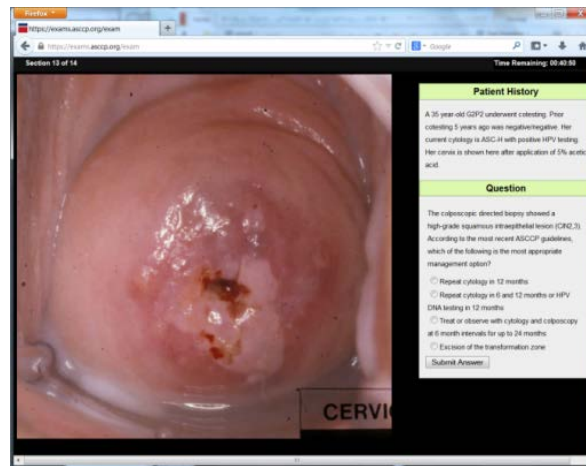
classifiers for automated classification of the epithelium into classes of normal or various grades of abnormal. In the latest experimental work, we achieved classification accuracy of 88 percent in the automatic labeling of the epithelium as normal or one of the three classes of abnormal, using the LDA classifier with 27 image features to label each of 61 epithelium images that a pathologist with expertise in the histology of the uterine cervix had labeled.

In FY2014, we installed and began the trial use of

technology developed in collaboration with Texas Tech University to view, manually segment, and then do tissue classification of regions in uterine cervix histology images through a Web-based system. This system includes the Intelligent Scissors algorithm (well known in the image-processing community) to assist and speed up the segmentation process and will take advantage of graphics processing unit (GPU) hardware for some of the image-processing computations, if a GPU is available on the user's local system.



LHNCBC automatically processed this image to find the center line (“medial axis”) for a segment of uterine cervix tissue, and then to divide the tissue into multiple segments for detection of cervix disease. (A collaboration among the Communications Engineering Branch, Missouri University of Science & Technology, and the University of Oklahoma Health Sciences Center.)



The American Society for Colposcopy and Cervical Pathology (ASCCP) continues to use LHNCBC's Teaching Tool software to administer online, nationwide proficiency exams in the field of colposcopy. This figure shows a representative question in how to manage a patient, given clinical history and a current colposcopy image of the uterine cervix.

Open-i: Biomedical Images Search Engine for Clinical Decision Support, Research, and Education



Two collections added to NLM’s biomedical image search engine *Open-i* allow scientists and the public to view images in PubMed Central articles related to their searches (upper left), drill down to details about images of interest (bottom left), and view related images in the collection of chest radiology reports provided by Indiana University (upper right) and anatomy illustrations provided by University of Southern California (bottom right).

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 LHCBC Clinical Information Systems effort, which exploits ongoing research in both natural language processing and content-based image retrieval by processing and indexing images based on both text and image features. We developed the *Open-i* system for finding images and figures in published literature or other sources. It 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. Users may search by text queries as well as by example images. They can filter images by type (e.g., X-ray, graph), filter journals by clinical specialty, and rank papers by clinical task (e.g., treatment). This ability assists clinical decision support, biomedical research, and education.

Open-i was released to the public in 2012 and is the first production-quality system of its kind in the biomedical domain. The system enables users to search and retrieve medical citations from 517,715 open-access

articles in PubMed Central® (PMC). In FY2014, we increased the collection of images in Open-*i* to 2.2 million images from PubMed Central articles (from 1.3 million in FY2013) and to 7,470 from Indiana University radiology reports.

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 ranked among the best.

The demand for Open-*i* services grew 26 percent in 2014. We upgraded the system architecture and the user interface to be consistent across desktop computers and mobile devices, and increased the number of images shown per page from 20 to 100.

The system is available 24/7, and it can handle more than 20,000 interactions per day in real time. We have also added an application programming interface (API) that provides batch-retrieval services for researchers who need access to images. For example, Pittsburgh researchers used the Open-*i* API to build a visual ontology for carcinoma biomarkers and add illustrations to their oral squamous cell carcinoma dataset. We capture the most popular searches and display the top images for the three most popular searches on the Open-*i* home page. This new feature allows users to quickly navigate to the results of popular searches that were trending in 2014, such as for Ebola, keloids, pulmonary tuberculosis, and lung tumor.

The evolution of Open-*i* continues to be supported by research in these key areas:

- Information retrieval and search engine optimization;
- Text summarization;
- Representing images with text strings;
- Combining global and local representations of image features;
- Improving methods for automatically segmenting multi-paneled illustrations into single images and partitioning their captions to correspond to those single images; and
- Improving methods for extracting pointers (arrows, arrowheads, symbols) within images to identify regions of interest that could be more relevant to a query than the entire image would be.

Computational Photography Project for Pill Identification (C3PI)

The Computational Photography Project for Pill Identification (C3PI) is developing an authoritative, comprehensive, public digital-image inventory of the nation’s commercial prescription, solid-dose medications. Working with expert consultants, we are 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 FY2014, we increased our online collection to more than 808,000 images of more than 3,000 pill (solid-dose pharmaceutical) samples 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 is to develop computer-based tools to help identify an unknown medication based on image-matching algorithms.

These images are freely available online directly via an API or through an interactive Web search via NLM’s Pillbox and RxNav sites. Through two online repositories—<http://rximage.nlm.nih.gov> (which links to RxNav) and <http://splimage.nlm.nih.gov>—we can distribute images of oral, solid-dose medications to the public and to pharmaceutical manufacturers, respectively.

In FY2014, we began a collaborative project with the Food and Drug Administration (FDA) to extract drug-drug interactions from drug labels to support FDA’s structured product labeling (SPL) indexing initiative.

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 source for the anatomical data needed to model human structure and physiology. 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 so the images retain their original resolution) format to licensees over the Internet at no cost and on DVDs for a duplication fee. Almost 44,000 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 FY2014, 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.

Insight Tool Kit

The Insight Toolkit (ITK) is a public, open-source algorithm library for segmenting and registering high-dimensional biomedical-image data. In FY2014, we released the official software version ITKv4.6.0, which includes a collection of more than 1,500 filters and algorithms for medical image processing. 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.

In FY2014, ITK's project manager, Terry Yoo, PhD, received the Hubert H. Humphrey Award for Service to America, "for leadership in open science through ITK as the public software standard for 3-dimensional biomedical-image analysis and for all the patients who have benefited."

We also updated SimpleITK (sITK) to broaden accessibility of ITK to the Python programming language community. Developed primarily by programmers at NLM, sITK is a simplified layer built on top of ITK. It is 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. ITK remains an essential part of the software infrastructure of many projects across and beyond NIH. Since its inception in 2004, the Harvard-led National Alliance of Medical Image Computing (NA-MIC), and NIH National Center for Biomedical Computing (NCBC), has adopted ITK and its software-engineering practices as part of its engineering infrastructure. Though the 10-year funding for the NCBCs is now complete, ITK remains integrated with 3D Slicer, the open-source software surgical-planning workstation developed by NA-MIC. 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 application programming interface (API) for integrating robotics, image guidance, image analysis, and surgical intervention. In FY2014, the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) awarded a Small Business Technology Transfer (STTR) grant to build a virtual-surgery system to treat children with craniosynostosis. The new project will build on the experience of the development of IGSTK, ITK, and 3D Slicer.

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 University of Utah's Scientific Computing and Imaging Institute's SCIRun, and through developing a new release of VolView, free software for medical volume image viewing and analysis.

Over the past year, we have seen innovation in and adoption of both ITK and sITK. The NLM software initiative continues to have impacts in research and education. At least three educational initiatives have used sITK to teach medical volume image viewing and analysis, including courses at Rensselaer Polytechnic Institute (RPI), the University of Iowa, and the Imperial College of London. In keeping with the principles of open science, the tutorial and educational materials are all made available without cost or license restrictions online.

3D Printing

NLM has been contributing to the NIH 3D Print Exchange, an online resource for searching, browsing, and downloading 3D printable files related to medicine and biosciences. This consortium of three NIH ICs—NIAID, NLM, and NICHD—received funding and backing from the HHS Ignite and HHS Ventures initiatives of the HHS IDEA Lab. The 3D Print Exchange officially launched at the White House Maker Faire in FY2014.

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 Toolkit (ITK) and Kitware's Visualization Toolkit (VTK) to automatically generate printable models from X-ray CT data. The project's data repository is officially online, and a national advisory board is assisting with its oversight, testing, and guidance.

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.

Natural Language Processing and Text Mining

Medical Article Records System (MARS)

NLM's flagship database, MEDLINE®, contains more than 21 million bibliographic records for articles from more than 5,600 biomedical journals. To meet the challenge of producing these citations in an affordable way, researchers at LHCNBC develop automated techniques to extract bibliographic data, such as abstract, author names, and affiliations, from both scanned-paper and online journals.

While the bulk of citations now comes to NLM directly from publishers (in XML format), nearly 707 journals provide citations in paper form only. These are

processed by the Medical Article Records System (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. The stages in this automated process are:

- Segmenting page images into text zones,
- Assigning content labels (title, author names, abstract, grant numbers, etc.) to the zones, and
- Pattern matching to identify the specific entities in each zone.

We have also developed a system (Publisher Data Review System, or PDRS) to supplement the citation data that publishers send in electronically (in XML form), since these often contain missing fields. Examples of such missing items required in MEDLINE citations are databank accession numbers (e.g., for items in GenBank and similar databases), 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. PDRS subsystems are based largely on machine-learning algorithms such as the support vector machine, or SVM.

To extend automated indexing data extraction to *all* online journals on publishers’ sites, including the journals 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 it retains articles (for processing) for only a very short time in random access memory (RAM). The IMPPOA system:

- 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, and
- 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 IMPPOA 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.

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 semiautomated 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 Unified Medical Language System (UMLS®) Metathesaurus, which are then restricted to Medical Subject Headings (MeSH®). The second approach uses the MeSH 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. Indexers consulted the MTI recommendations for 63 percent of the articles in FY2014, compared with 59 percent in FY2013. MTI also provides indexers with an option to select MeSH heading-subheading pairings for some of the MTI recommendations based on our subheading attachment software. The MTI developers have also created special filtering for 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 has been designated as the first-line indexer for those journals. As a “first-line” indexer, MTI indexing is subject to human review. As reported in the Library Operations chapter, the number of journals for which MTI is the first-line indexer continues to increase, saving time and money.

Our collaboration with LO and experience gained by participating in the BioASQ challenges—for biomedical semantic indexing and question answering (<http://www.bioasq.org/>)—have helped increase performance of the MTI system.

In FY2014, MTI provided the primary baseline for the second year of the international BioASQ challenge. 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. MetaMap, a critical component of the MTI system, is also used worldwide in bioinformatics research. In FY2014, users downloaded about 2,100 copies of MetaMap and 1,200 copies of the Java API and *Unstructured Information Management Architecture* (UIMA) wrapper.

Digital Preservation Research (DPR)

The long-term preservation of documents in electronic form, both born-digital and scanned from paper, is a mandate for NLM, as it is for other major libraries and archives. The goal of the LHCBC DPR project is to

investigate and implement techniques for these key preservation functions: automatically extracting metadata to enable future access to the documents, ingesting the documents and metadata into a storage system, and conducting knowledge discovery on 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, Resource Description Framework (RDF)) while incorporating in-house-developed modules that implement the preservation functions listed above.

In 2014, we completed the preservation of a historic medico-legal collection of early 20th-century court documents acquired 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. NLM curators have been using SPER to preserve the FDA documents, numbering 67,000 in total. In 2014, we completed the processing of the fourth and final set of court case summaries. This set comprised 30,700 notices of judgment, published between 1940 and 1966 and related solely to adulterated or misbranded foods. The documents were added to the archive, and, along with metadata, they are publicly accessible at an NLM Web site and are being used by researchers. 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 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; and
- 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.

RIDeM/InfoBot

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

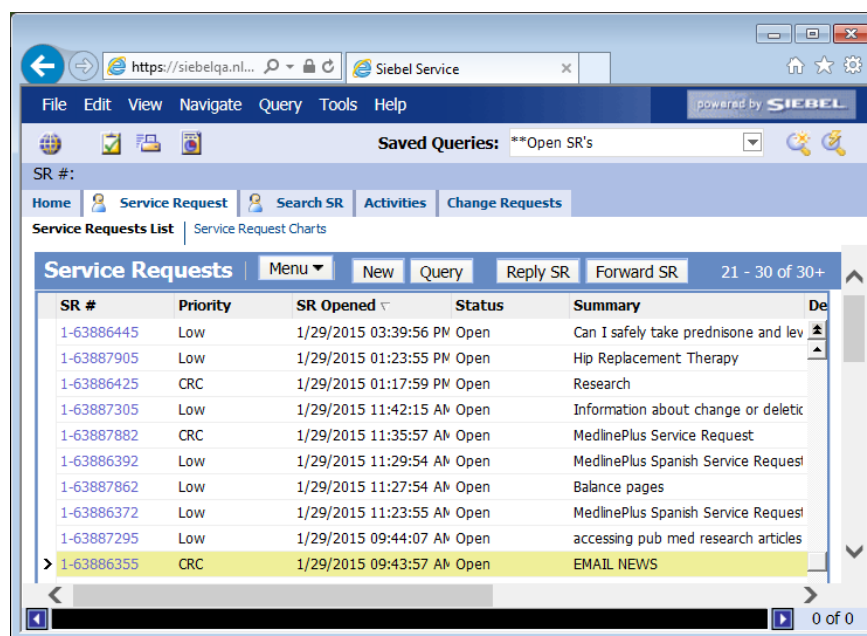
The related InfoBot project enables a clinical institution to automatically augment a patient’s electronic medical record (EMR) with pertinent information from NLM and other information resources. The RIDeM API developed for InfoBot allows just-in-time access to patient-specific information to be integrated 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 2014, the tab was accessed 615 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 100,000 requests a year. In FY2013, we developed and evaluated a prototype CHIQA system. 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.

In May 2014, the PubMed Corrections Assistant was integrated into NLM’s customer service workflow. It automatically prepares stock replies for NLM customers. Fifty percent of replies generated by the CHIQA PubMed Corrections Assistant do not need editing by customer services staff. Research into improving question classification and recognition is ongoing.



The Consumer Health Information and Question Answering system has automatically prepared 4,000 responses for review by customer services staff.

De-identification Tools

De-identification allows the clinical research community to study clinical data without breaching patient privacy. We are developing a clinical text de-identification system called NLM-Scrubber to automatically remove patient identifiers from narrative clinical reports. The provisions of the Privacy Rule of the Health Insurance Portability and Accountability Act require the removal of 18 individually identifiable information elements that could be used to identify an individual or that individual's relatives, employers, or household members. We completed a version of the software system that de-identifies clinical narrative text in a form of electronic messaging known as Health Level Seven (HL7) version 2. It can use information embedded in various HL7 fields as well as externally provided information, such as the list of names of the healthcare providers at NIH.

We are using another tool we developed—the Visual Tagging Tool (VTT)—to develop a gold standard collection for developing and testing NLM-Scrubber. By the end of 2014, we amassed a collection of 23,446 clinical reports of 8,150 patients, in which human reviewers manually labeled every piece of individually identifiable information. We are working on deploying the NLM-Scrubber to the NIH Clinical Research Information System

(CRIS), which is NIH's clinical research repository. All clinical narrative reports generated by NIH clinicians and researchers will be de-identified using NLM-Scrubber. We also shared the VTT with the natural language processing community to be used for other types of lexical tagging and text annotation.

Librarian Infobutton Tailoring Environment (LITE)

Infobuttons are links from one information system to another that anticipate users' information needs, take them to appropriate resources, and help them retrieve relevant information (<http://www.infobuttons.org>). They are mostly found in clinical information systems (such as electronic health records (EHRs) and personal health records (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,” that can be a national resource for electronic health record developers and users (<http://www.openinfobutton.org>). 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 knowledge bases to do their customization work. The knowledge bases are very institution-specific. We developed the LITE, a user-friendly tool that an institution’s medical librarians can use and that 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>). Which is now responsible for its maintenance as an open source tool. A user-evaluation project is now under way in collaboration with Ohio University.

Lexical 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. The 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.

The 2014 release of the SPECIALIST lexicon contains 484,628 records representing more than 896,000 forms, an increase of 7,771 records from the 2013 release. Many of the new records are derived from de-identified clinical records. Others were identified using an n-gram database of MEDLINE.

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 drug,

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.

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 more than 70 million SemRep predications extracted from all MEDLINE citations; the database 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.

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 FY2014, we performed extensive work related to NLM-developed and -supported clinical vocabulary and messaging standards that are crucial to the success of the Federal Health Information Technology (HIT) goals. We provided comments and assistance to the Office of the National Coordinator for HIT (ONC), Centers for Medicare and Medicaid Services (CMS), and FDA, often on extremely short deadlines, about meaningful-use regulations for electronic health record vocabulary and clinical-data-standards requirements, with special focus on improving the fit of quality measures to the realities of EHRs and avoiding efficiency losses in clinical office practices. LHCBC’s Director served on the Trans-NIH Outcomes-Effectiveness Research Interest Group, the National Center for Advancing Translational Sciences (NCATS) Global Rare Disease Registry Steering Committee and Common Data Elements Working Group, the National Children’s Study Information Management System Review, and the American Medical Informatics Association (AMIA) EHR-2020 Task Force. He was also selected to serve on the Content Standards Workgroup of the HIT Standards Committee—the Federal Advisory Committee charged with making recommendations to ONC on standards, implementation specifications, and certification criteria for the electronic exchange and use of health information.

In FY2014, we participated in an effort to align SMART, CIMI, and FHIR behind NLM-developed and -supported clinical vocabularies LOINC (Logical

Observation Identifiers, Names, and Codes—for medical tests, measurements, and observations), SNOMED® (Systematized Nomenclature of Medicine—Clinical Terms), RxNorm (for medications), and UCUM (for computable units of measure). Our SNOMED CT® mapping projects, described below, support Federal Meaningful Use EHR requirements to use SNOMED CT for patient problems.

Medical Ontology Research

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

During FY2014, we investigated the coverage and representation of human phenotypes in standard terminologies and developed unsupervised machine-learning methods for extracting drug-drug interactions from the structured product labels. We also developed a prototype version of MeSH in RDF (Resource Description Framework) for use in the Semantic Web as a Linked Open Data resource. We continued supporting the NLM Value Set Authority Center, focusing on the quality of drug value sets in clinical quality measures. Finally, in collaboration with the Center for Drug Evaluation and Research at the Food and Drug Administration (FDA), we developed methods for extracting adverse drug events from MEDLINE indexing.

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). SNOMED CT is the designated US standard for coding the problem list and many other EHR fields.

We analyzed problem-list vocabularies and their usage frequencies in 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. In FY2014, we maintained and updated the CORE Problem List Subset of SNOMED CT, which now contains 6,100 concepts and has been used in EHRs, terminology research, and quality assurance activities. In addition to the standalone downloadable file, the CORE Problem List Subset is also available as a reference set (refset) in the international release of SNOMED CT and in the UMLS as a specific content view. It is updated four times a year to synchronize with changes in SNOMED CT and the UMLS.

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.

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 United States, ICD-9-CM (the ninth version of “Clinical Modification”) is also used for reimbursement (to be replaced by ICD-10-CM in October 2015). 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 better suited for that purpose, and its use is required as part of the meaningful-use regulations.

To assist 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 that allows users to encode patient problems in SNOMED CT terms and then generate the appropriate ICD-10-CM codes in real time for billing or other purposes. In FY2014, we expanded the coverage of the SCT-to-ICD-10-CM map to 54,000 concepts (from 35,000 in FY2013). The map has been used by 1,200 UMLS licensees. To demonstrate the use of the map, we developed the I-MAGIC (Interactive Map-Assisted Generation of ICD Codes) demo tool. In FY2014, I-Magic attracted 22,281 visitors, up from 15,271 in FY2013.

For an international project, in collaboration with the IHTSDO and the World Health Organization (WHO), we helped develop an analogous rule-based map from SNOMED CT to ICD-10. We also adapted our I-MAGIC tool to showcase this map. 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 from 3,000 ICD-9-CM procedure codes.

RxTerms

RxTerms is a free, user-friendly, efficient interface terminology for drugs 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, at least one EHR from a major medical institution in Boston, and by some proprietary-drug-software vendors.

We update RxTerms monthly to synchronize with the full release of RxNorm. In addition to the standalone download, RxTerms is available through the RxNorm API.

RxNav

Released in September 2004, RxNav was first developed as a graphical interface browser to the RxNorm database and was primarily designed for displaying relations among drug entities. In addition to the RxNav browser, we later created SOAP-based and RESTful application programming interfaces (APIs) to let users integrate RxNorm functions into their applications. We also integrated additional drug information sources, including RxTerms, the National Drug File-Reference Terminology (NDF-RT), and pill images.

During FY2014, we developed RxClass—an application (and companion API) for linking RxNorm drugs to drug classes from the Anatomical Therapeutic Chemical (ATC) classification of drugs, from the Medical Subject Headings (MeSH), and from DailyMed. RxNav, RxClass, and the APIs received a combined total of about 230 million queries during FY2014 (a significant increase from 70 million queries in FY2013). Users include clinical and academic institutions, as well as pharmacy management companies, health insurance companies, EHR vendors, and drug information providers. Developers of mobile apps have also started to integrate our APIs into their applications.

LOINC® Standards for Identifying Clinical Observations and Orders

Within medical record systems, patient summaries, and reports to public health organizations, Federal Meaningful Use EHR regulations require that lab result messages sent to ordering clinicians use LOINC (Logical Observation Identifiers, Names, and Codes). In FY2014, 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. The FDA Mini-Sentinel program is adopting LOINC for its lab codes, and the National Patient-Centered Clinical Research Network (PCORnet) is encouraging the use of LOINC for all variables. A hospital in Buenos Aires is now using LOINC, Infobutton, and NLM's MedlinePlus Connect to deliver information about lab tests to patients' personal health records.

By the end of FY2014, LOINC had more than 32,000 users in 162 countries and had been translated into 15 languages, including new translations into Dutch, French, and Portuguese. The Regenstrief Institute and the International Health Terminology Standards Development Organisation (IHTSDO) signed a long-term agreement to begin collaborative work on linking their global healthcare terminologies: LOINC and SNOMED Clinical Terms.

We worked with Regenstrief and the LOINC Committee to create more than 1,200 new LOINC terms

for both laboratory and clinical variables, and the LOINC database now contains nearly 74,000 terms. We released new terms for toxicology, chemistry, hematology, microbiology, and clinical survey assessment instruments (e.g., AHRQ's (Agency for Healthcare Research and Quality's) Healthcare Event Reporting Form).

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 common data 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 these institutes:

- 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 Diseases Research, to revise the 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 LOINC codes for the neurological quality-of-life (Neuro-QOL) measures done a few years ago.

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.

Big Data to Knowledge (BD2K): Using Large Clinical Databases to Assess the Associations among Patient Factors, Medication Usage, and Patient Outcomes

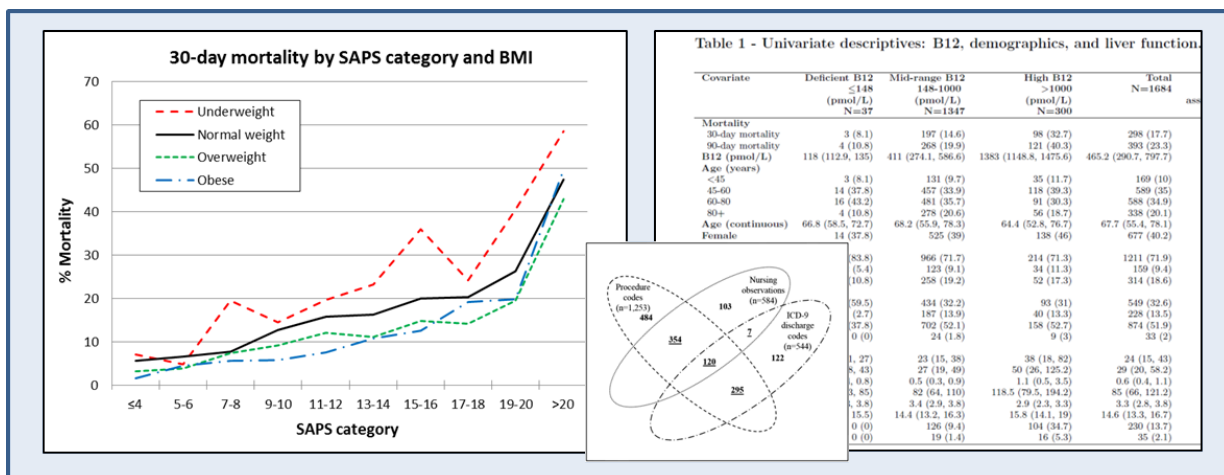
These projects support NIH's Big Data to Knowledge (BD2K) initiative and the NIH Director's research priorities. We used large clinical databases to predict patient outcomes from patient factors, including

medication usage. We combined structured data and information-retrieval techniques and developed and implemented algorithms. Ongoing studies illustrate the potential for Big Data to generate knowledge vital to improving clinical care.

We have been working with MIMIC II for a number of years. MIMIC II is a database organized by MIT and funded by NIBIB and NIGMS that collects almost all medical record data from ICU admissions at one large hospital, de-identifies the data, and organizes it in a usable database. It includes all ICU nursing variables: vital signs, orders, and information about local interventions, medications prescribed and dispensed, nursing notes, and radiology reports. It has physician admission and progress notes for the NICU babies, and otherwise only discharge summaries. It also includes all laboratory results produced during the hospital stay at which the ICU admission occurred. The current version of MIMIC II carries nearly 280 million rows of data about 32,536 patients (8,087 are neonates and the rest adults), plus waveform files. In the earliest versions these data were provided as local codes that were very duplicative. For example, there were more than 30 distinct codes for serum glucose. Here at LHNCBC, we translated all of their many different local codes to universal codes—LOINC for lab test and clinical measures, RxNorm for drugs and SNOMED CT for problems. We returned this data to MIT

and they incorporated it into their distributed version so that all of the MIMIC II users could take advantage of these codings. MIMIC II also carries some information collected from outside of the ICU, including vital status (alive or dead) obtained by the originating hospital before de-identification via linkage to the Social Security death tapes.

We have completed and published a number of studies on the MIMIC II data. Investigators are now looking at the relation between use of statin medications by ICU patients and the effect on mortality during sepsis, because the anti-inflammatory effects of statins are postulated to be protective in such situations. One of our fellows looked at the concordance between a well-known formula based on laboratory test for predicting Disseminated Intravascular Coagulation (DIC) in ICU patients and the diagnoses and therapeutic actions taken by providers. Using neonatal intensive care unit data for 75 infants with necrotizing enterocolitis (NEC) and 189 controls thoroughly matched on gestational age, gender and being small for gestational age, we have established the role of feeding in NEC. Addition of cow milk-based products to breast milk significantly increased the risk of NEC for premature infants. Infants receiving combination feeds had more than triple the risk of developing NEC as those on exclusively breast milk feeds, and those receiving only formula had seven times the risk.



In 2014, we continued researching secondary use of large collections of clinical data from the MIMIC II Intensive Care Unit database.

We obtained access to the CMS Enclave that provides analytic (but not direct) access to the new and improved version of Medicare and Medicaid data and patient assessments, called “CMS Chronic Conditions Data Warehouse (CCW).” The CCW is a research database designed to make Medicare, Medicaid, Assessments, and Part D Prescription Drug Event data more readily available to answer research questions. Beneficiary matching, deduplication, and merging of the files in preparation for a study—which in the past had to be done by each researcher—has already been done by CMS. With the Enclave, researchers can analyze the CCW data while the data remains in place on CMS’s computer, via SQL queries and SAS, but cannot see (or download) any patient level data or aggregations of very small samples. Because of this extra layer of privacy protection, permissions are easier to get, and for other reasons the cost is much less. We obtained permission to analyze a large subset of the database with a special focus on drugs, treatments, and Part D. We have begun projects looking at the use of influenza vaccine based on charges for the dispensing, the effect of better adherence to the current recommendations for when to get the flu shot (considering the actual peaks of incidence), and at the overall rate of flu shots dispensed compared to patients’ reports of usage in the Medicare survey.

Newborn Screening Coding and Terminology Guide

We’ve collaborated with many federal, state, and other agencies to standardize the variables used in newborn screening (NBS) by using national coding standards required by Meaningful Use. 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 FY2014, 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.

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

We perform extensive research on 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 comprehensive information about registered clinical research studies readily available to the public. Each month during the past year, it received more than 135 million page views and hosted about 960,000 unique visitors. Nearly 16,200 study sponsors, including the federal government, pharmaceutical and device companies, and academic and international organizations, submitted data to ClinicalTrials.gov through the Web-based Protocol Registration and Results System (PRS). At the end of FY2014, the site had more than 176,200 research studies, conducted in all 50 states and in more than 180 countries. Approximately one-third of the studies were still open to recruitment. For the remaining two-thirds, the recruitment phase was either over or the study had been completed. Summary-results tables describing primary and secondary outcomes, adverse events, and characteristics of the participants were posted for more than 14,600 of the registered studies.

In FY2014, new registrations of clinical trials were submitted at an average rate of 460 records per week, an increase of 15 percent from FY2013. The average rate of new results submissions was about 90 per week, an increase of 29 percent from FY2013. We attribute the continued growth in the use of ClinicalTrials.gov to US laws that require registering and reporting the summary results of clinical trials, as well as international recognition of the scientific and ethical importance of registering 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.

ClinicalTrials.gov staff continued to educate the public about the most recent federal law concerning clinical trials and results submission, Section 801 of the Food and Drug Administration Amendments Act of 2007 (FDAAA 801). We continued to work with the NIH Office of the Director, other NIH Institutes and Centers, and the FDA on a Notice of Proposed Rulemaking (NPRM) that elucidates the requirements of FDAAA 801 as well as key implementation issues. We also participated in the development of an NIH draft policy on the registration and results submission of all NIH-funded clinical trials. (Both the NPRM and the NIH draft policy were issued in November 2014 for public comment.)

Since FY2013, we have been evaluating and enhancing the PRS site. Sponsors and investigators use the site to submit, update, and maintain study registration and summary-results information. On the basis of direct user feedback, findings from usability studies, and recommendations from usability experts, we continued implementing user interface improvements intended to streamline the data entry process in FY2014. We also continued providing targeted education and outreach on the results database and submission requirements through developing additional help materials, presenting at conferences, participating in working groups, and publishing in journals.

ClinicalTrials.gov research projects and publications in FY2014 included:

- Identifying reporting discrepancies between results of trials posted on ClinicalTrials.gov and corresponding findings published in the peer-reviewed literature.
- Using ClinicalTrials.gov data to estimate the gap in US federal human research oversight regulations.
- Characterizing prematurely terminated trials registered on ClinicalTrials.gov and determining whether primary outcome data are available from such trials in the ClinicalTrials.gov results database and the published literature.

Genetics Home Reference (GHR)

The 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 by using the rich technical data from the Human Genome Project and other genomic research. At the end of FY2014, GHR offered user-friendly summaries of 2,376 genetics topics, including more than 1,000 genetic conditions, 1,240 genes, all the human chromosomes, and mitochondrial DNA. GHR also offers an online handbook called Help Me Understand Genetics, which provides an illustrated introduction to fundamental topics in human genetics, including mutations, inheritance, genetic testing, gene therapy, and genomic research.

In the past year, the GHR team expanded the Web site's genetics content for consumers, adding 252 new genetics summaries and two new Help Me Understand Genetics pages about genetic susceptibility and informed consent. We also reviewed and updated 157 existing summaries. The team hosted a virtual intern from the University of Maryland, who helped us add and update links to consumer resources throughout the Help Me Understand Genetics handbook. In FY2014, the GHR Web site averaged 16.5 million page views per month and 61,500 visitors per day (an increase of 22 percent and 45 percent, respectively, from FY2013).

In support of SNOMED CT's initiative to include more rare and genetic diseases, we submitted more than a dozen updates for existing terms and synonymy in SNOMED CT. We also continued to integrate 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 and, beginning in April 2014, from ICD-10-CM (International Classification of Diseases, 10th Revision, Clinical Modification). GHR topics, each of which can map to multiple SNOMED CT codes, mapped to more than 2,800 SNOMED CT codes at the end of FY2014.

This year, GHR began a new contract with Genetic Alliance, an umbrella organization for condition-specific genetics interest groups. The new contract continues our three-year collaboration with Genetic Alliance to update existing GHR Web site content and track new research developments about particular genetic conditions. We updated about 60 existing GHR topics through this initiative in FY2014. In another collaboration, the GHR team consulted with researchers at the National Human Genome Research Institute (NHGRI) to develop content for use in a study involving genetic counseling for carriers of autosomal recessive genetic diseases.

Beginning in July 2014, GHR implemented a customer satisfaction survey in partnership with ForeSee, Inc. The initial survey, which ran for one year, will provide information that the team can use to improve the content, look, and navigation of the GHR Web site. We also added "Add This" buttons to each page of the Web site this year, allowing users to share GHR content via e-mail, Facebook, Twitter, and other social media. Additionally, we performed outreach activities to increase public awareness about GHR. For example, 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 and clinical and molecular fellows at NHGRI.

Profiles in Science Digital Library

The Profiles in Science® Web site showcases digital reproductions of items selected from: the personal manuscript collections of 36 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 and 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 future 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.

In FY2014, we added two new collections. Sir William Osler (1849–1919) revolutionized American medical education with the clinical internship program he instituted at the Johns Hopkins School of Medicine. Senior students worked in hospital wards full time to “learn medicine at the bedside.” The Osler addition was made possible through collaboration with McGill University and Johns Hopkins University. Mike Gorman (1913–1989) was a well-known journalist, author, publicist, and crusader for health policy reform who won a Lasker Award in 1948 for his newspaper exposés of state mental hospital conditions in Oklahoma.

The Web site averages more than 105,000 unique visitors each month, including people seeking an authoritative source of information about current events. The Web site experienced a spike in interest during the HHS and White House celebrations of the 50th anniversary of the Surgeon General’s 1964 Report on Smoking and Health.

Turning the Pages (TTP)

The goal of this project is to give laypeople 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 for online versions. 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 10,000 unique visitors a month and 326,000 page views a month. The iPad app is also popular: in FY2014, 4,536 people downloaded books with the app for the first time (new users), and more than 77,130 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 12 rare books. In FY2014, we added two books to the iPad version and one book to the kiosk and Web versions.

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 include the following:

- 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 enable users to use two or more fingers to perform tasks such as zooming in and out.

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 research findings to help determine the best course of treatment for a patient. A “consensus abstracts” element provides rapid review of multiple publications via smartphones.

To evaluate the usefulness of abstracts in clinical decision making, randomized controlled trials using simulated clinical cases 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. These studies demonstrated the usefulness of the app for clinical decision making.

The PubMed4Hh app is available for iOS (iPhone/iPad) and Android users. In FY2014, the iOS app was downloaded 240,000 times by users in the United States (45 percent of downloads) and elsewhere (55 percent). Queries from smartphone apps now account for 60 percent of all queries. The total number of searches has tripled since before the smartphone app was introduced.

Disaster Information Management: Lost Person Finder

NLM’s increasing interest in mitigating the effects of wide-area disasters has 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, location, 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 can be reported to PL by hospital counselors, relief workers, and the public. PL users can report or search for the missing via computer or mobile apps. We made various codebase enhancements, including beginning the process of converting the site to a responsive design framework. During FY2014, we enhanced imaging algorithms and software to reunite families in the wake of disasters, and we added visual search capabilities through use of FaceMatch services.

Deployments

In 2014, PL was deployed for three events: Typhoon Neoguri (July), the Jammu-Kashmir Floods (September), and Typhoon Hagupit (December). In the case of the Jammu-Kashmir Floods in Pakistan and India, more than 10,000 missing-person records were posted to PL (many through Google's Person Finder, which is interoperable with PL). Of these records, about 2,000 had photos of missing people. Additionally, NLM's Specialized Information Services (SIS) used People Locator to demonstrate several disaster-related tools produced by the NLM in the disaster informatics program at the NLM Biomedical Informatics Course held in Young Harris, Georgia (September). Social networking project accounts were reviewed and updated for more effective communications. A new social media process and strategy were added to include preparedness messaging for proactive notifications before and during potential disaster events.

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.

For a regional Maryland disaster drill conducted in FY2014 and covering five counties, Suburban Hospital used TriagePic to capture photos and metadata of incoming patients that were searched later for family and friends. We also demonstrated to drill participants the visual search method outlined in the Face-Matching Research section. For community-wide reporting, the ReUnite® app is used and was updated to be compliant with a major mobile operating system release (iOS 8.0+).

Face-Matching Research

Our goal 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 has 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 2014, we made considerable progress in face-descriptor matching and retrieval. We researched and developed a new Rotation-and-Scale Invariant Line-based Color-aware (RSILC) descriptor and confirmed (on the available annotated face datasets) that it is more accurate in matching line-rich objects (e.g., faces) than any individual legacy descriptor used by FaceMatch (FM), but not more accurate than the ensemble of the legacy descriptors.

In FY2014, our FaceMatch's visual query system performed for our CalTech Faces benchmark dataset at a 98 percent top-1 hit rate, which means that in 98 percent of visual queries, we expect the correct record to appear as the first listing in the results set, with subsecond query turnaround on typical People Locator datasets. We increased its top-1 hit rate from 95 percent in FY2013. Our R&D about automatic face-skin-tone clustering has not shown conclusive results yet, and more experiments are needed.

In an effort to make the face retrieval more robust and efficient, we currently invest in R&D efforts for the person-in-the-loop approach to both face localization and face matching. We collect the problematic images (with false-positive and negative-face detections and misclassifications), analyze our current algorithms' limitations, and investigate incremental learning approaches to skin mapping, face detection, and face matching.

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.

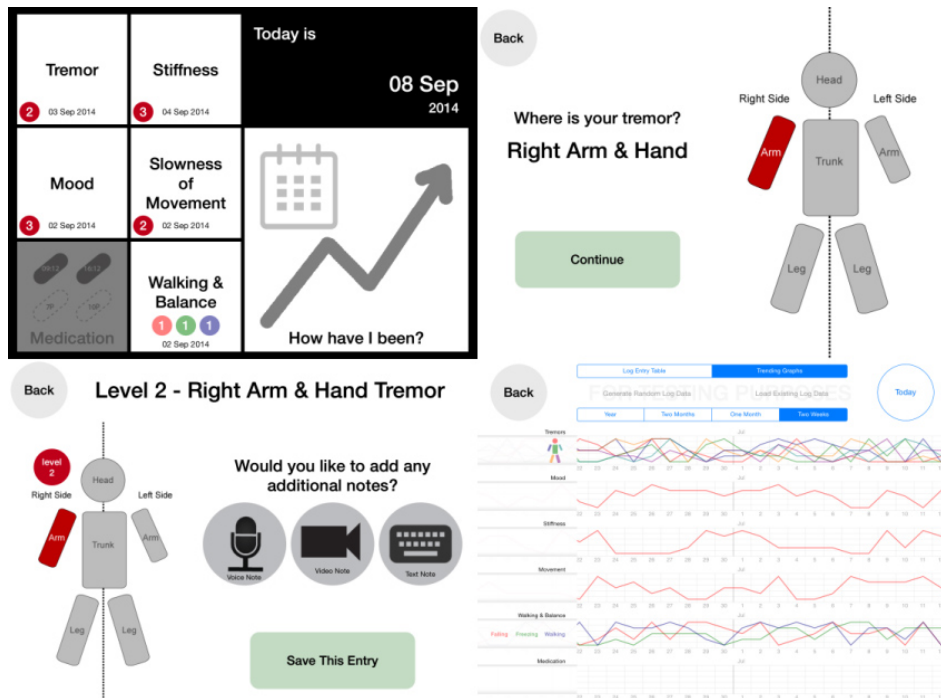
Movement Disorders Video Database/MDmedia

This ongoing LHCBC 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 (PD) by patients and their caregivers. In FY2014, in collaboration with the National Institute of Neurological

Disorders and Stroke (NINDS) and the NIH Movement Disorders Clinic, LHCBC developed a prototype mobile application (for iPad), based on the Movement Disorders Video Database (MDVD), to aid patients with PD and their caregivers in tracking patient symptoms and medication side effects. The *Movement Disorders Journal (MD Journal)* app provides an easy-to-use mobile platform to track speech, dyskinesia, tremor, mood, slowness of movement, and walking and balance on a daily basis. In addition, an interactive log has been developed within the *MD Journal* app that allows patients to track their dyskinesia status every 30 minutes over a 24-hour period.

This 24-hour log could replace the paper-based traditional PD diary that patients currently use. The long-term goal is to provide patients with PD and their caregivers with a reliable medical history to be shared with the patients' physicians through a cloud-based or other data sharing system and to be incorporated into the patients' electronic health records (EHRs).

In collaboration with the NIH Movement Disorders Clinic, the *MD Journal* app will be tested by a small group of patients with PD over a 6-month period to determine usability and usefulness of this app in helping to manage patients' disease.



APDB designed a prototype mobile record system which allows patients to track symptoms, medication side effects, moods, and other parameters including personal video recording. NINDS is using this system to test the research hypothesis that patients with Parkinson's Disease can actively manage their care with this app, resulting in improved overall disease management and health outcome. Initial evaluation of mobile app is underway at the NINDS clinic with movement disorders patients and caregivers.

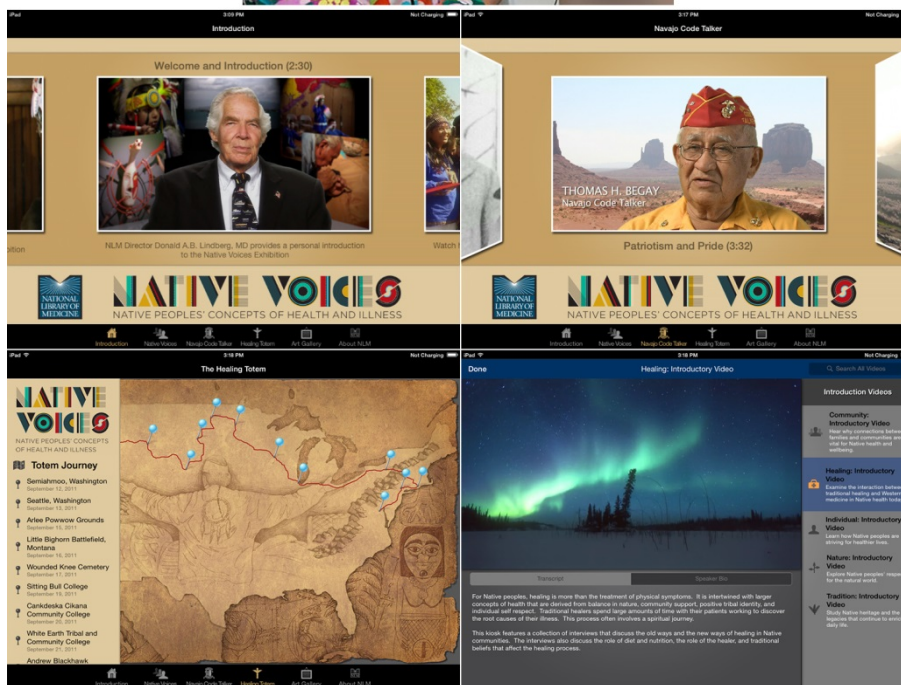
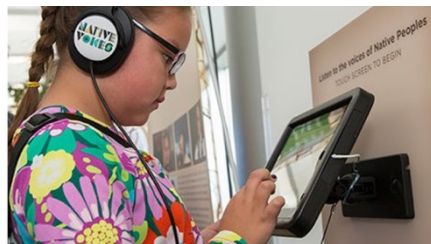
Native Voice Mobile Exhibition Planning, Development, and Deployment

In FY2014, we completed the pilot phase of the Native Voices Mobile Adaptation traveling exhibit, which included successful opening programs and presentations. 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, APDB provided extensive programmatic, technical and logistical planning for all venues, including:

the National Congress of the American Indian, Dena'ina Convention Center, Anchorage, AK; The Native American Heritage Center, Anchorage, AK; Southcentral Foundation Native Primary Care Center, Anchorage, AK, Bentuh Nuutah Valley Native Primary Care Center, Wasilla, AK; University of Alaska Anchorage, WWAMI Program, AK; The Queen's Medical Center, Honolulu, HI; John A. Burns School of Medicine (JABSOM), Honolulu, HI; The Hamilton Library at the University of Hawai'i at Manoa; Artesian Art Gallery, Chickasaw Nation, Sulphur, OK; Texas Medical Center Library, Houston, TX; University of Washington, Seattle.

The traveling exhibition pilot survey instrument was developed for implementation on an iPad Mini, which has been traveling with the mobile exhibition to all venues. Preliminary data has reported outstanding results, with high marks for content, interface, video materials, and

overall design. In partnership with the NLM's Regional Medical Libraries, NLM will implement a second phase deployment of the Native Voices traveling exhibition in FY2015.



(Top) A girl at the exhibition kiosk uses the Native Voices iPad app. Screenshots from the Native Voices iPad app, showing: (1) NLM Director Donald A.B. Lindberg, MD presents a personal introduction to the Native Voices Exhibition, (2) Navajo Code Talker Thomas H. Begay discusses patriotism and pride, (3) map of the Healing Totem Journey, and (4) Introductory Video about the theme of Native Healing.

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 (HD) video over Internet Protocol (IP), we continued to monitor the HD open-source work of video conferencing tool (VIC) developers on H.264 compression.

Our team also continued investigating 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 by NLM, are either too complicated or unstable, or have licensing terms that are not cost-effective given the needs of the Library. We continue to believe that there is no current compelling rationale to migrate from the H.323 standard videoconferencing appliances used for NLM program support. The team also initiated a review of mobile videoconferencing apps. We continued to collaborate with the Rochester Institute of Technology (RIT) and the University of Puerto Rico Medical Campus to test open-source software and commercial cloud technologies for compressed HD videoconferencing based on the H.264 video standard. A manuscript about our cloud findings was accepted for publication, and we are summarizing the

results of our mobile app study in a paper we plan to publish in 2015.

We decided to work exclusively with UltraGrid technology for uncompressed high-definition 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, and
- the research team at Masaryk is very active and shares our research goals and interests.

In 2014, the team initiated the testing of UltraGrid's extended 4K compressed and uncompressed capabilities, which were added by Masaryk. A clinical trial of uncompressed video was completed at the Medical University of South Carolina (MUSC), and the equipment there was returned. The trial's purpose is to study the use of uncompressed video as a diagnostic tool. Investigators selected teledermatology as a research focus because previous research showed that it is particularly difficult to use standard-definition video to perform remote dermatological exams. We are now reviewing the data we received and following up on some missing data.

We continued to work with SIS on a distance-education outreach program for minority high school students and with the NIH Library to offer distance-education training on the National Center for Biotechnology Information (NCBI) database and other bioinformatics resources. In FY2014, 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. In addition, we began working with the NIH Library to offer distance-learning courses at NIH Institutes located outside Bethesda, especially NIDA in Baltimore and NIEHS in Research Triangle Park, NC. A high school in Puerto Rico was added to the minority student outreach program with SIS. Video archives of these and certain other programs are available on the Collab Web server. In 2014, all programs were converted from a Real Media format to one that can be accessed without plug-ins by any HTML 5—or later—browser. Finally, the team supported a two-day workshop linking, by videoconference, US medical librarians brought to NLM and academicians in Tanzania planning to start health information science programs. The workshop was sponsored by NLM's International Program.

Training and Education at LHCBC

LHCBC 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 LHCBC research projects under the guidance of 16 LHCBC research staff mentors. Participants also make presentations, write manuscripts, attend and present at professional conferences, and may publish in professional journals.

During FY2014, 40 trainees (including 19 postdoctoral fellows) from 18 states and 7 foreign countries received training and conducted research at LHCBC in a wide range of disciplines, including:

- 3D image processing,
- Big Data to Knowledge (BD2K) based on large biomedical datasets,
- Biomedical ontology and terminology,
- Clinical health information question answering (CHIQA) systems,
- Collaboration tools,
- Content-based information retrieval,
- De-identification of medical records,
- Evidence-based medicine systems,
- Information retrieval,
- Literature-based discovery,
- mHealth, image, text, and document processing,
- Natural language processing,
- Personal health records,
- Pill identification,
- Semantics Web research, and
- Disaster management.

We emphasize diversity by participating in programs for minority students, including the summer internship programs of the Hispanic Association of Colleges and Universities and the National Association for Equal Opportunity in Higher Education.

The MITP sponsors the Clinical Informatics Postdoctoral Fellowship Program, funded by LHCBC, to attract young physicians to NIH to pursue research in informatics. This program is run jointly with the Clinical Center to bring postdoctoral fellows to labs throughout NIH. 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 gives trainees from NLM-funded medical informatics programs an opportunity to learn about NLM programs and current LHCBC research. The rotation includes a series of lectures showcasing research conducted at NLM and provides an opportunity for trainees to work closely with established scientists and fellows from other NLM-funded programs.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

David Lipman, MD
Director

Highlights of the Year

Collaboration to Identify Outbreaks of Foodborne Illnesses

In FY2014, NCBI made significant progress in a multi-agency project to identify outbreaks of foodborne pathogens. The initiative, which involves the Food and Drug Administration (FDA), the Centers for Disease Control and Prevention (CDC), the Department of Agriculture (USDA) and state health laboratories, is using the advanced technology of whole genome sequencing (WGS) to detect foodborne disease outbreaks with greater specificity and improved speed than previous methods.

As part of this project, CDC, FDA, USDA, and state public health labs use WGS on isolates from food and environmental samples and compare those to the clinical isolates from sick patients. NCBI collects the sequencing data, assembles the sequences, and compares them to other sequences in the pathogen database. When a match is found, the geographical scope of the outbreak can be rapidly and precisely determined and steps taken to prevent further cases by locating the source of the pathogen. This methodology is now being deployed to field laboratories and has proven valuable in quickly tracing recent outbreaks of Salmonella and Listeria.

PMC Becomes Foundation for Federal Agency Public Access Plans

On February 22, 2013 the White House Office of Science and Technology Policy (OSTP) issued a memorandum calling on federal agencies with over \$100 million in research expenditures to develop plans to support increased access by the public to their published research results. This requirement closely followed the NIH Public Access Policy mandated by the Congress in 2008. A number of agencies from inside HHS and across the Government have expressed interest in using the NCBI's PubMed Central (PMC) infrastructure for their public access programs.

Among HHS agencies, CDC, FDA, AHRQ, and ASPR will utilize PMC, thereby increasing the number and breadth of articles in PMC to include more research topics in comparative effectiveness, emergency preparedness, public health, environmental health, and toxicological research.

Mission

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. In November 2013, NCBI celebrated its 25th anniversary, commemorated by a statement read into the Congressional Record by US Senator Ben Cardin, who commented that NCBI has emerged as "an impressive national resource 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, as have genome-wide studies with genotypic and phenotypic 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, chemistry, 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 and services 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 and their protein translations. 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 US Patent and Trademark Office for sequences from issued patents (PAT), ensure that the collection contains all available relevant data. GenBank data is available at no cost through FTP and a wide range of Web-based retrieval and analysis services.

GenBank is divided into separate divisions based on organism taxonomy and sequence data collection methods. Twelve taxonomic divisions (BCT, ENV, INV, MAM, PHG, PLN, PRI, ROD, SYN, UNA, VRL, VRT) contain sequences for over 300,000 formally described species. High-throughput sequencing divisions include GSS, EST, HTC, STS, and TSA. The Whole Genome Shotgun (WGS) division includes contigs from WGS projects.

Together, the WGS and non-WGS portions of GenBank contain over 403 million sequence records and

over 973 billion basepairs. The database continues to grow rapidly, with the number of bases doubling approximately every 18 months. The traditional nucleotide sequences division increased to 174 million records in FY2014, from 167 million records in FY2013. The WGS division grew to over 189 million records, up from 124 million records in FY2013. In FY2014, the 200th version of GenBank was released.

GenBank records are created by indexers with specialized training in molecular biology who apply rigorous quality controls. NCBI taxonomists consult on organism classification and, as a final step, senior NCBI scientists review the records for biological accuracy. In order to simplify access to, and improve the quality of, the enormous amounts of data stored in GenBank, NCBI is continually developing new tools and enhancing existing products and methods. 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. 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, genome consortia, and sequencing centers.

GenBank data submission can be accomplished via two resources: BankIt and Sequin. The BankIt submission tool allows the author to submit sequences within a series of online forms 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. Recent improvements to both tools have been made to enhance usability.

Genome Information Resources

NCBI plays a key role within 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. NCBI maintains an ever-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, dbSNP, reference sequences (RefSeq), CCDS, dbGaP, Gene, Probe, UniGene, HomoloGene, dbVar, ClinVar, Genome, 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 representing 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. They provide a stable reference for gene identification and characterization, mutation analysis, expression studies, polymorphism discovery, and comparative analysis.

Manual curation by NCBI staff is provided for different aspects of data, depending on the section of the RefSeq project. For example, protein names are curated for Microbial RefSeq proteins. Curation of the transcript and protein sequence, gene structure, and functional information is provided for a subset of mammals, other vertebrates, invertebrates, and plant genomes. The latter category of curation relies on a combination of automated and manual sequence analysis, publication review, and collaboration. In the past year, the total number of organisms supported by the sequence curation group increased by 134 percent to 220 distinct organisms. Curation and automated analysis of individual genes, alternate splice variants, and individual nucleotide mismatches (relative to the reference human genome sequence) resulted in over 47,634 new or updated RefSeq protein-coding, non-coding, and pseudogene records. During FY2014, 31,432 new or updated records for human were added, bringing the total number of curated RefSeq human records to over 70,000. An additional 33,000 uncurated human RefSeq transcripts were generated by NCBI's annotation pipeline, and the best supported of these are added to the curated dataset over time.

At the end of FY2014, the entire RefSeq collection contained over 45 million proteins for over 49,000 organisms, representing a 30 percent increase in the number of proteins from last year and a 56 percent increase in the number of organisms. The Fungal section of RefSeq realized the largest growth in the number of organisms (257 percent), followed by bacterial and archaeal organisms (60 percent), plants (50 percent) and vertebrates (38 percent). The greatest growth in the number of proteins occurred for organisms with large genome sizes, including vertebrates (180 percent growth), followed by the plants (111 percent) and fungi (85 percent); proteins for bacterial and archaeal organisms grew by 24 percent.

During FY2014, RefSeq curators also helped define metadata and input sequence data to initiate two process flows—one propagates annotation from GenBank genome submissions to RefSeq, and the other is an evidence-based, *de novo* whole genome annotation pipeline. NCBI staff identify GenBank genomes that are in scope for either process, gather information about available relevant transcript data, and initiate the appropriate process flow. In addition, curators define custom BLAST databases, add or update organism pages

for the Genome resource, and perform QA review of annotation result presentation prior to public release in the Map Viewer resource.

In FY2014, NCBI staff processed 54 eukaryotic genomes through the newly re-designed annotation propagation pipeline, and 87 organisms through NCBI's eukaryotic genome annotation pipeline (a 74 percent increase). Additionally, the RefSeq prokaryotic genome annotation pipeline processed 13,938 genomes during the year; this included an initiative to re-annotate almost all existing RefSeq bacterial genomes in order to provide whole genome annotation using a single method, thus making the annotation more comparable across taxa.

The NCBI eukaryotic annotation propagation pipeline was recently re-developed to use an agile and robust process flow. RefSeq curation staff made a significant contribution to the development process by reviewing output, identifying data issues, and communicating the nature of the issue to development staff. This pipeline re-development is approximately 75 percent complete.

RefSeq curation staff serve as a point of contact for over 80 research communities; they provide education about genomics resources and RefSeq data, advocate for data submissions, and respond to community requests for curatorial improvements for transcript and protein representation of 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 30,499 CCDS IDs for human and 23,880 CCDS IDs for mouse (five percent and three percent growth, respectively). 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 approved by members of an international collaboration. During FY2014, over 256 existing CCDS entries were subject to collaborative review and 285 explanatory public notes were added to the database. The CCDS collaboration also worked during the year 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 complex transcribed regions to define genic regions, including non-coding RNA loci and pseudogenes.

NCBI's Gene database provides a unified query environment about genes and other loci defined by sequence and/or map location. Gene 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. The database is heavily used and manages information for

more than 12 million genes from more than 12,000 taxa. About 200,000 full records are retrieved each weekday from about 45,000 user sessions. In order to maintain a high standard of service, Gene dataflow 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. Gene now displays gene location on both current and prior genome assemblies, including both human GRCh38 and GRCh37.p13 assemblies. Also, support for Ensembl and INSDC annotation tracks has been expanded, helping users see similarities and differences in annotation, and convert between datasets. Numerous improvements were made to the Gene Web interface, and for reporting information about HIV interactions with human genes. A review of the Gene resource that describes improvements made in the last four years was recently published.

The largest user group for Gene is those accessing data on human, mouse, and other medically important taxa. Therefore, Gene staff is heavily involved in ensuring high-quality annotation of genes on genomic sequences for these taxa. In FY2014, the number of species annotated with NCBI's eukaryotic genome annotation pipeline expanded to 158 taxa, many of which do not have annotation available in any other public resource. During the year NCBI annotated a major update to the human reference assembly (GRCh38) and made the data available in Gene, RefSeq, and other resources less than two months after initial release of the assembly, allowing users to rapidly start taking advantage of the new information.

RefSeqGene, a subset of the RefSeq database, defines genomic sequences to be used as reference standards for well-characterized genes. The representation of human genes as RefSeqGene records continued to grow during the year, to over 5,100 records. Activity within the Locus Reference Genomic (LRG) (subproject) collaboration also increased. The aim of the project is to create an LRG for every locus with clinical implications. Presently, over 790 RefSeqGenes are assigned LRG accessions.

The primary tool for visualization of assembled genomes is NCBI's MapViewer. Genes 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 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 for specific areas of interest. In FY2014, several back-end updates to MapViewer were implemented to improve the efficiency of data loading and indexing.

Genome Decoration Page uses a different approach, allowing users to place their own annotation

data on graphical ideogram representations of the genomes of more than 20 organisms. From this page, users can also access the Bulk Sequence-Cytogenetic Conversion Service. This related tool, which also can be accessed via an API, allows users to get cytogenetic locations for features or sequence locations, or to conversely get sequence locations for cytogenetic locations.

The related NCBI Remap tool allows users to remap coordinates of genes, SNPs, and other markers from one genome assembly to another. This tool accepts a variety of genome annotation formats and uses the alignments of one assembly to project annotation features on another 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. It also supports mapping of features annotated on Locus Reference Genomic (LRG) sequences, used in the reporting of genomic variants. In FY2014, NCBI implemented front-end updates to facilitate use of the Web interface and back-end work to provide improved remapping of variation data. Review of logging data showed an increase in usage over the previous year.

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 FY2014, the GRC released a coordinate changing update to the human reference genome assembly, the culmination of approximately four years of curation. The GRC also released a coordinate changing update to the zebrafish reference assembly and a non-coordinate changing patch update to the mouse assembly. Concomitant with this work, the GRC focused its attention on tasks associated with providing support for curation of additional genomes and organisms. NCBI work involved updates to the underlying database and associated Webpages, collaboration with external groups sequencing other human and non-human genomes, providing assembly QA, and contributing analyses for publications.

This year NCBI also continued its collaboration with the Maize Genome Database to pursue their use of GRC curation tools for updates to the maize genome reference assembly; efforts included assisting with sequence analysis and training in assembly curation.

Comparative Genome Resources

The Genome database provides sequence and map information for a wide variety of organisms including eukaryotes, archaea, viruses, and bacteria. The Genome site recently 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 pathogen genome sequencing, metagenomes, transcriptome sequencing, genotyping, variation and epigenomics. The resource provides tracking of several data elements, including detailed information about a project's scope and the granting agency. In FY2014 over 10,000 bioprojects were submitted via the Web-based submission system and an additional 25,000 were created through other submission streams, including GEO and other non-Web-based submission systems. A BioProject chapter is included in the *NCBI Handbook, 2nd Ed.*

NCBI's Assembly database tracks the set of sequences that comprise a genome assembly and the structure of the assembly, such as contigs, scaffolds, chromosomes and gaps. The database provides a stable assembly accession number with versions that are incremented when submitting groups update their assemblies. The Assembly database provides metadata about assemblies, such as assembly names and alternate names, simple statistical reports of the assembly (type and number of contigs, scaffolds, N50s), and a historical view of updates. It also tracks the relationship between an assembly submitted to the INSDC and the assembly represented in the NCBI RefSeq project.

In FY2014 the Assembly database added 14,361 new genome assemblies (13,717 prokaryotic assemblies and 644 eukaryotic assemblies) from 3,443 different species. This addition increased the number of assemblies in the database by 93 percent and the number of species represented by 49 percent. Three updates to the Assembly Web resource were performed in FY2014. These updates: enhanced the search functionality, making it easier to find genome assemblies of interest; added a link to a BLAST Web page preconfigured to search against the genomic sequences in the assembly; and enriched the Assembly details page with more metadata that help differentiate between multiple genome assemblies for a particular species.

NCBI's ftp (File Transfer Protocol) site is an important resource for a multitude of users. In FY2014, the genome area of the ftp site underwent a major revision to expand available content and facilitate data access through an organized directory hierarchy. The initial ftp release provided a standard set of data files for over 45,000 genomic assemblies, organized in three new directories within the genome area. The common data unit within the directories is a subdirectory corresponding to a record in the Assembly database and given a name consisting of the assembly accession followed by the name of the assembly. This single process data flow will result in comprehensive provisioning of data and a consistent core set of files for all organisms and assemblies.

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 and can be viewed with the Taxonomy database browser. The 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. Taxonomy report pages now include organism "type material" information (specimen or strain). During FY2014, the Taxonomy database reached a milestone of 300,000 individual records of species with formal scientific names.

The Sequence Read Archive (SRA) is NIH's primary archive of high-throughput sequencing data. The repository accepts various formats of high-throughput sequencing data to meet the demands of current science and technology. In FY2014, SRA reached two petabytes of sequence data. 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 is available for download. The source code for each SRA release also is available for download.

Clone DB is a resource that integrates information about clones and genomic libraries, including sequence data, map positions and distributor information. In FY2014, curation efforts led to the creation of new records for 35 organisms, a nearly 33 percent increase in represented taxa. Additionally, clones were placed on assemblies for 15 organisms, including multiple human assemblies. These placements are provided to MapViewer and are visible on individual clone record pages, where the data may be mined to identify clones containing features of interest or guide genome assembly improvements.

The Gene Expression Omnibus, or GEO, is a public repository for large-scale functional genomic data generated by microarray- and sequencing-based technologies that consists of two divisions. GEO DataSets is a study-level database and GEO Profiles is a gene-level database. In fiscal year 2014, GEO processed over 10,600 new studies, representing 21 percent growth in the database. GEO development improvements this year included further streamlining and modifying submission templates and documentation in order to encourage provision of richly described data. Also, existing records were reviewed and updated to conform to updated vocabularies. A new pipeline was developed to submit processed next-generation sequence data to NCBI's vertical database (VDB) and Track Management System (TMS), thus enabling users to visualize thousands of next-generation sequence tracks on NCBI's Sequence Viewer.

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 provides displays (“genome tracks”) of the raw data mapped to genomic coordinates. The database contains over 4,800 studies, covering more than 1,400 samples from five well-studied species.

NCBI’s BioSample database contains descriptions of biological source materials used in experimental assays. In FY2014, the BioSample database received over 600,000 new samples, increasing the database by 30 percent. During the year, many improvements were made to the BioSample Submission Portal, including the addition of new BioSample Packages, enhanced validation and auto-fixing of provided data, and harmonizing BioSample attributes, all helping to enhance content quality. Other developments included implementing faceted searching, switching to a Service Orientated Architecture, supporting XML Schema 2.0, and generating ASN1 reports for GenBank and other databases.

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 currently contains information on over 15 million probes.

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 updated releases were provided to the public in FY2014, supplying users with a number of new features including: export of alignments (BAM/cSRA) to tab-delimited format, flexible broadcasting between bio trees, improved searching in Test view, the addition of a history list to the Open dialog, adaptive features that display when zooming in Multiple Alignment view, support for Linux OpenSUSE 13.1 operation system, a GFF format reader, new Tree render with improved speed, redesign of the data import to support molecule id mapping from local to NCBI accessions using genome assemblies, implementation of filtering/search in Open View dialog, and the addition to Graphical Sequence view of visualization for map data records representing restriction maps or other non-sequence based maps.

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 complement other information-rich views. In FY2014, four

NCBI Sequence Viewer versions were released with new features including: new options for visualization of graphs, improved fonts, a “Preferences” option for setting display parameters, compacted tooltips, the ability to select/deselect multiple tracks in configuration dialog, additional support for an NCBI authentication protocol for restricted datasets, improved error processing for inaccessible tracks, improved large files upload, FASTA downloads for mRNA/proteins/exons, multi-track upload, improved display of long track titles, improved embedding documentation, initial support for touch-screen user interface, integration of track upload with MyNCBI accounts, and an improved set of readers (GFF, VCF, BED).

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 65,900 new influenza virus sequences were added to the Influenza Sequence Database in FY2014. Contributors included the NIAID Influenza Genome Sequencing Project, the US Air Force, and other institutions worldwide.

Variation Data Resources

The database of Short Genetic Variations (dbSNP) catalogs common genetic sequence variation. dbSNP contains over 356 million submissions of human genome data that have been processed and reduced to a set of 113 million non-redundant refSNP clusters. Approximately 350 other organisms are represented in the SNP database, with one billion submissions curated to 500 million refSNP clusters. In addition to a traditional Entrez and dbSNP Web search, human SNP data are also accessible on NCBI’s new Variation Viewer and Variation Reporter tools, which draw on dbSNP, dbVar, ClinVar, and NCBI’s own human genomic annotation.

The database of genomic structural variation (dbVar) contains data on variant DNA from many organisms and related clinical data. 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. The dbVar Genome Browser allows users to browse multiple studies at once and search for any feature annotated on the genome.

ClinVar archives interpretations of the relationship between human variations and observed health status. A participant in the NIH ClinGen project, ClinVar is a key component of the subset of NCBI resources that are focused on medical genetics. Interpretations for about 80,000 new variants were submitted to ClinVar in FY2014. More than 128,000 variations from dbSNP and dbVar are linked to ClinVar reference accessions with clinical assertions. ClinVar is experiencing steady usage growth, receiving about 2,500 requests each day.

The 1000 Genomes Project is an international consortium that aims to build a comprehensive map of human genetic variation. In FY2014, the project began its final data release, representing 2,600 individuals originating from 25 populations around the world. In addition to cataloging variation data, 1000 Genomes has served as 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 SVs 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 results graphic shown within the browser is based on NCBI's Sequence Viewer graphical tool. Below the graphical view, a table of genotypes organized by the 1000 Genomes Project is shown from which users can drill down to show individual-level genotypes. The 1000 Genomes browser underwent back-end updates in FY2014 to incorporate and take advantage of updates that were occurring to the NCBI Track Management System (TMS), including track search, and to support display of data from the project's final data release.

Genome browsers are available for a number of other NCBI resources and FY2014 saw the development of several new browsers, as well as improvements to existing browsers. Development also continued for the Track Management Service (TMS) and User Uploaded Data (UUD) resources, both of which support browser-related data.

First implemented in FY2013, the NCBI Track Management System (TMS) is a 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. This year, using TMS, several thousand new data tracks were loaded into Gene, Clone DB and GRC applications. In FY2014, a second TMS instance was established to support management of tracks originating from dbGaP, a controlled-access database. This work was integral to the development of the dbGaP Data Browser (DDB). In addition, TMS began indexing of tracks, which provides the necessary infrastructure to develop search interfaces. A limited-scale TMS search was implemented in an updated version of the 1000 Genomes project browser and DDB; an interface supporting a comprehensive TMS search is planned for FY2015.

User Uploaded Data (UUD) is a centralized resource that ensures various NCBI tools provide consistent handling of data uploaded by users. Data uploaded via UUD is managed by TMS. In FY2014, updates to the UUD design continued, including implementation of a test suite, improved handling of large files, statistics collection, and support for permanent storage. By the close of FY2014, a UUD interface ("Your

Data") was found in all TMS-enabled browsers, allowing users to upload their own data for display in the context of NCBI-provided data.

The Genome Data Viewer (GDV) is a generic browser application developed in FY2014 that allows visualization of data within a genomic context, using an NCBI TMS, SRA or SADB accession. A resource-agnostic browser, GDV is currently used by GEO to provide a graphical display of data associated with specified experiments or samples. Adoption of GDV for display of data in other NCBI resources (e.g., genotypes and data in SRA) is expected to follow.

The Variation Viewer, which enables browsing of human variation data in a genomic context, was launched in FY2014. This new browser serves as an integration point for variants cataloged in multiple NCBI resources, including dbSNP, dbVar and ClinVar. It supports variants annotated on the current and previous versions of the human reference genome assembly. Users can query the browser by location, phenotype, gene or variant IDs. Further discovery is accomplished in the Variant Table, which includes information on clinical significance from ClinVar, allele frequencies and publications. A comprehensive set of table filters also allows users to drill down into results to find their variants of interest.

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 International HapMap Project samples NA12878 and NA19240 from several clinical testing and research laboratories. NCBI and the CDC are also collaborating with the Genome in a Bottle Consortium, which has generated high-quality variant calls for NA12878. All of the data is available from NCBI's GeT-RM Web site. The GeT-RM browser displays the variants in genomic context along with genotype information, and allows data for a specific region to be downloaded. Full datasets can also be downloaded from the GeT-RM ftp site.

Other Resources

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. At the end of FY2014, 417 laboratories from 39 countries had submitted data to GTR on 23,276 tests for 5,031 tested conditions. Compared to FY2013, the number

of labs has increased by 13 percent, the number of tests by 91 percent, and tested conditions by 38 percent. This growth was supported by enhancements to data submission formats in FY2013 and further enhancements in FY2014 that enabled submission of all GTR fields via a comprehensive Excel template. Submission enhancements and staff support also enabled submission of multiplex panels. At the close of the fiscal year GTR included more than 1,141 registered tests that evaluate five or more genes or chromosomal locations, representing a growth rate of 280 percent. To facilitate retrieval of GTR's rich data, an advanced search for tests was implemented on the public site, and comprehensive extraction of GTR data in an XML format was made available on the ftp site. During FY2014 GTR maintained a pattern of steady, recurring use similar to that of other regularly used NCBI Web resources.

MedGen, a companion resource to GTR and ClinVar, is NCBI's portal to information about human disorders and 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.

MedGen has integrated 268,018 concepts from seven vocabularies, including two which are of particular value to the genetics community and are not yet provided by the UMLS Metathesaurus: Human Phenotype Ontology and Orphanet. 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 programmatically via E-utilities.

During FY2014, MedGen maintained a pattern of steady, recurring use. While MedGen has not yet achieved wide usage, it has become well-recognized by several important stakeholders including the NHGRI-funded Clinical Genome Resource (ClinGen). For example, MedGen has provided solutions to several ClinGen workgroups, such as the Phenotype workgroup, Education workgroup, and Gene Curation workgroup.

Chemical Information

PubChem is organized as three inter-linked databases: Substance, Compound, and BioAssay. Together, they form an extensive resource for millions of chemical substances. In the 10 years since its launch in 2004, PubChem has grown to a key information resource for a number of research fields including cheminformatics, chemical biology, medicinal chemistry, drug discovery, and RNAi research.

PubChem reached significant growth milestones in 2014 including: over 180 million contributed substance records, over 63 million unique compound records, 1.1 million bioassay records, and 296 data contributors. The PubChem BioAssay database now contains more than 228 million bioactivity test results from: over three million tested substances, 723,000 RNAi reagents, 9.7 thousand protein targets, and 30,000 gene targets. PubChem routinely receives more than one million requests per day from an estimated 140,000 users.

During the year there were 50 new data contributors, broadening the scope of PubChem's content and expanding public access to these data. The number of substance and compound records increased by 48 percent and 31 percent, respectively, and the number of bioassays grew by 50 percent. Importantly, the U.S. Food and Drug Administration (FDA)'s Structured Product Labeling (SPL) group contributed information on marketed pharmaceuticals. In addition, the Environmental Protection Agency (EPA) Substance Registry Services—which is the authoritative resource for basic information about chemicals, biological organisms, and other substances of interest to EPA—deposited information on substances that are tracked or regulated by EPA. The PubChem Upload system now can provide data contributors with dynamically created private URLs that allow them to share their on-hold data with collaborators or reviewers before the data becomes publicly available on PubChem. This new feature helps to support staging of data in PubChem prior to publication or patent filing.

In early 2014, PubChem released PubChemRDF, which encodes PubChem information using the Resource Description Framework (RDF). PubChemRDF allows users to utilize schema-less databases (such as a triplestore or a graph database) to import, query, and analyze PubChem data on local computing resources using semantic Web technologies. It also harnesses ontological frameworks to help facilitate PubChem data sharing, analysis, and integration with resources external to NCBI and across scientific domains. PubChemRDF data can be accessed through a REST interface or bulk downloaded from the PubChem FTP site.

During the year, PubChem revamped the Compound Summary page, which recaps all information known about each chemical in the Compound database. Considerable emphasis was given to making the Compound Summary page load faster by reducing the amount of data required to display the page and by

improving the PubChem computational infrastructure. The new interface uses a responsive design approach; it is optimized for both touch- and mouse-based interfaces and automatically adjusts to the available screen size, making it user friendly for desktops and mobile devices. The Compound Summary page contains new annotations from various sources, for example, Unique Ingredient Identifiers (UNII) and pharmacological classifications from FDA, Anatomical Therapeutic Codes (ATC) and International Nonproprietary Names (INNs) from the World Health Organization (WHO), and environmental health information from the EPA.

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.

In FY2014 a new Identical Protein Report display option was added to protein reports. The new display provides a list, in tabular format, of all other identical proteins, including those submitted as translations to GenBank, as well as RefSeq, outside protein databases and patented protein records. Links to the CDS sequence in Nucleotide are available for each protein.

Molecular Modeling Database (MMDB)

The Molecular Modeling DataBase (MMDB) is an archive of experimentally determined three-dimensional structures of biopolymers and biopolymer complexes that mirrors the content of the Protein Data Bank (PDB) and is updated weekly following the release of novel structures by PDB. MMDB integrates three-dimensional structure data with other data collections at NCBI, such as molecular sequences, literature, and chemicals. It provides seeds for novel protein domain families, as defined for the Conserved Domain Database (CDD). MMDB assigns uniform secondary structure and structure domains for polypeptide structures and uses these to pre-compute structural similarities between protein chains or domains via the VAST algorithm (Vector Alignment Search Tool). VAST+ a recent extension to VAST, pre-computes structural similarities between molecular complexes. The CBLAST service links a large fraction of protein sequences tracked by NCBI's Entrez system to the three-dimensional structures of homologous proteins.

In FY2014, MMDB staff prepared the computational pipeline that is used for importing and processing molecular structure data to work with a revised data format provided by the PDB (the mmCIF standard), as some macromolecular structure entries will no longer be available in the traditional PDB format, due to its inherent limitations. This will simplify the processing of large macromolecular structures.

Conserved Domain Database (CDD)

The Conserved Domain Database (CDD) is a resource for the annotation of protein sequences with the footprints of functional motifs conserved in molecular evolution. Two CDD releases were processed in FY2014. CDD curators added a total of 2,729 curated or validated models to the in-house tracking database during the year.

CDD version 3.11 marked the initial public release of annotation computed with a revised RPS-BLAST search algorithm that employs composition-corrected scoring. Corresponding search models (position-specific scoring matrices), which support new options available in RPS-BLAST versions 2.2.28 and up, also were made available.

The display style for drawing domain models in CD-Search, Batch CD-Search, and CDART has been revised. The display style is now uniform among those tools, with a given domain model rendered in the same shape and color by all three tools. Additionally, a new display option, "Standard Results," is available in CD-Search and Batch CD-Search, and shows the top-scoring domain model from each source database.

The CDD curation group continued to work on models that represent repetitive structural motifs that are too small to be classified as globular domains. The first set of such models has been added to the public sequence annotation resource. CDD was presented at various scientific meetings in FY2014.

The CDD group continued work with Dr. Andrew Neuwald from the University of Maryland on automating protein domain classification via sampling subsets from a multiple sequence alignment that conserves subfamily-characteristic sequence signatures. Work in FY2014 also resulted in a production version of the subfamily sampler that will be used for a data production study. Research undertaken by a summer student compared manually curated sequence alignments to alignments generated by several fully automated methods.

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 FY2014, the BioSystems database added over 106,000 new records from source databases, an increase of 20 percent.

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, review recent search results, and 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. In FY2014, organism BLAST pages were updated to use top-level RefSeq genomic records rather than scaffold records. This change, using hits on chromosome coordinates, is more useful for public reporting and for comparative analysis.

Stand-alone BLAST provides a platform for users wanting to perform searches locally. The latest version released this fiscal year was 2.2.29. In FY2014, NCBI began providing a BLAST installation at Amazon Web Services in an effort to deliver services to users wanting to use new cloud technologies. By accessing an Amazon Machine Image (AMI), users can run stand-alone searches with the BLAST+ applications with a simplified BLAST search page. In July, a Webinar entitled "BLAST in the Cloud" was provided to assist users with the new service. The Webinar is available on NCBI's YouTube page. NCBI also released a YouTube video on how to create custom databases for BLAST.

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, which 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 520 studies since 2007, 103 of which were released in FY2014. In addition to new studies, 135 previously released studies were updated with additional data (via new versions) in 2014. The new studies are comprised of phenotype and genotype data from 214,343 additional study participants, bringing the total number of study participants represented in dbGaP to 868,829.

An online registration system was created in FY2014, making it easier and more time efficient to register studies, submit data and access data. Tutorials for using the online forms are available on NCBI's YouTube site.

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), and dbGaP staff is actively working on establishing links to MedGen records using MeSH terms. 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 a total of 868,829 research participants.
- Over 162,110 individual-level phenotype and exposure variables.
- 3,725 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 study participant DNA sequence. These potentially reveal systematic and heritable genetic differences between affected and unaffected individuals.
- 3,828 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. As of the end of FY2014, 4,965 research projects from 3,006 Primary Investigators have been approved for access to dbGaP data. 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 FY2014, there was significant progress in the development of the dbGaP Data Browser (DDB). This browser was developed with input from the NHGRI-funded ClinGen project and is intended to help authorized users to view selected genomic regions across selected individuals, enabling discovery and improving efficiency of data downloads. The interface is modeled after the 1000 Genomes project browser and provides displays for variants, genotypes and supporting sequences reads. The content represents a combination of data from the dbGaP controlled access and public databases. The display of such mixed data required significant infrastructure development this year in the form of new browser and TMS components, as well sustained collaboration with the controlled access and Sequence Viewer teams.

Also new this year, the dbGaP team has begun to develop and test Genetic Relationship And Fingerprinting (GRAF), a fast and reliable computer algorithm and software package that can be used to determine the genetic relationships within a large array of study participants. These algorithms and values will also be beneficial to the users of dbGaP data sets, as well as other researchers.

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.

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.

In FY2014, two new options for data retrieval were added to E-Utils. ESearch offers a sort parameter function using the display settings menu options in a Web-based search. Also, outputs are available for JSON format.

Entrez Direct, a new software suite that enables users to use the UNIX command line to directly access NCBI databases, was released this year. The UNIX executables call E-Utilities directly and provide a variety of post-processing functions. The software is available via FTP and documentation is available on the NCBI Web site. All E-Utility programs are described in the Entrez Programming Utilities Help Manual.

Literature Information Resources

PubMed

PubMed provides Web-based access to citations and abstracts for 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 24 million citations from more than 37,000 journals, some dating back to the 1800s.

PubMed is continually updated and enhanced for better functionality and results. For example, in FY2014 PubMed began updating citations seven days a week, rather than five. Also, a group of PubMed subject filter strategies were revised to improve retrieval.

PubMed Mobile provides a simple Web interface for mobile devices. In FY2014, PubMed Mobile was updated with a new homepage and query box. Also, PM Commons indicators and "Related searches" have been added to the summary display. The abstract display now includes collaborator names and multiple author affiliations as well as search links on names.

My NCBI is an NCBI service that allows users to save searches and results, with added features such as automatic updates and e-mailed results. SciENcv is a My NCBI feature that helps users create an online professional profile and make it public to share with others. SciENcv complements the My Bibliography service, where users can save a record of their own citations. Changes this year include the ability to create multiple profiles, download profiles in various formats, add delegates to a profile, and add a mini profile to link to PubMed Commons.

PubMed Commons

PubMed Commons provides a forum for sharing information and perspectives about biomedical publications in PubMed. The project was developed in response to interest from the scientific community and has been operating as an open pilot since December 2013. This year, an external working group was established to advise the project.

PubMed Commons enables authors of PubMed-indexed publications to post relevant comments to any

PubMed record. Comments are posted immediately but are regularly monitored for adherence to guidelines. The comments appear below abstracts and can be viewed by anyone using PubMed. Comments are contributed under a worldwide, royalty-free, non-exclusive, perpetual Creative Commons license that permits sharing and reuse. Journal clubs may also apply to join PM Commons.

At the end of FY2014, there were 7,652 PubMed Commons members, and 1,934 live comments on 1,649 publications. Outreach efforts include a PubMed Commons Blog and a Twitter account with over 2,000 followers. New developments this year included a computational approach to use ratings and other information to increase or decrease prominence of comments, including highlighting comments at the PubMed Commons and the PubMed homepages, and refining the guidelines and formal system for monitoring and moderating comments. New features for users included a joining wizard, with major expansion of email addresses from PubMed authors to enable self-joining. In addition, the PubMed Commons team began providing a personal alerting service for authors of articles receiving comments, as well as tailored alerts on request to NIH ICs.

PubMed Health

The PubMed Health collection includes systematic reviews of clinical effectiveness research published in the last ten years. The collection grew to about 35,000 reviews at the end of FY2014. PubMed Health has extensive coverage of reliable systematic reviews, many in full text. The collection includes reviews from leading health technology assessment agencies (such as the US Agency for Healthcare Research and Quality), the Cochrane Collaboration, and those incorporated in the Database of Reviews of Effects (DARE) from the Centre for Reviews and Dissemination (CRD).

This year, the Swedish Council on Health Technology Assessment (SBU) became a systematic review partner, and the Canadian Agency for Drugs and Technology in Health (CADTH) started contributing content. The National Heart, Lung, and Blood Institute (NHLBI) became an information partner, contributing consumer information. Trials cited in many PubMed Health systematic reviews now link to their entries in PubMed, enabling PubMed users to go straight from a reports on a clinical trial to systematic reviews that have considered the trial (around 180,000 trials). A portlet in PubMed alerts users to these systematic reviews. In addition, My NCBI functionality for saving searches and creating alerts was added to PubMed Health.

A variety of curated information support tools were released, built around a growing site-wide medical glossary that had more than 1,200 entries at the end of the year. The PubMed Health team is curating these medical terms from dictionaries and glossaries produced by information partners and National Institutes of Health (NIH) sources. Some terms are also selected from

Wikipedia and Wiktionary, an outcome of PubMed Health's collaboration with WikiProject Medicine. Terms are supported by a feed of the latest systematic reviews and consumer resources for that subject, as well as identification of equivalent terms in US and British English, synonyms, and relationships between terms.

The PubMed Health blog keeps visitors informed about new features and material on the site. PubMed Health's Facebook followers grew from just over 1,000 to more than 8,300 during the year. The PubMed Health Twitter account grew to just short of 10,000 followers, and the Google+ account reached more than 3,000 followers.

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 2014, there were over 3.2 million articles in PMC. Average weekday usage levels were around 900,000 unique users and over two million articles retrieved.

At the end of FY2014 PMC had over 1,800 participating journals that deposit either all of their content or the subset of NIH-funded articles they publish. For publication year 2013, PMC has a little over 325,000 articles, about a third of which are associated with NIH funding (extramural research awards or intramural staff authors). About 40 percent of the NIH-funded articles came from journals that deposit final published articles directly into PMC; the remainder came as author manuscripts via the NIH Manuscript Submission System.

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 FY2014, the number of organizations participating in LinkOut increased to over 3,900. LinkOut participants include 3,100 libraries, over 530 full-text providers, and 270 providers of non-bibliographic resources, such as biological and chemical databases. Participation in Outside Tool, a service linking to external resources, increased to over 1,200 institutions. LinkOut users can now link to 60 million Entrez records, including links to the full text of 60 percent of PubMed records from over 11,000 journals. Usage of LinkOut resources reached over 40 million hits per month, and about 1.6 million hits per weekday.

In FY2014, all LinkOut utilities and programs were updated to handle Unicode to reflect the multilingual nature of the links. The quality assurance process is continually fine-tuned, focusing on automatic link checking to reduce broken links and improve user experience. Link submission in CSV format was also added to help providers to submit links.

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, resources, 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 120 investigators, staff scientists, research fellows, postdoctoral fellows, and students.

CBB is carrying out basic research on over 24 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 180 publications this year, with more in press. CBB scientists gave numerous presentations and posters at scientific meetings. Presentations were also given to visiting delegations from US and foreign institutions, oversight groups, and steering committees. CBB hosts many guest speakers and shares information about research projects at its weekly seminar 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 43rd meeting was held in December 2014. The research programs of three Senior Investigators were reviewed at the meeting.

Bioinformatics Training and Support

Outreach and Education

NCBI's outreach and public services component is essential in ensuring the user community is aware of NCBI services and is able to make effective use of those services. The audience for NCBI databases is very broad: not only molecular biologists and health professionals, but 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 fourteen articles on NCBI resources in the 2014 Database Issue of *Nucleic Acids Research*.

Educational materials published online are continually improved and updated with over 40 fact sheets and instructional sheets available through the education Web page. Written by NCBI staff and available on the Bookshelf, the *NCBI Handbook* is a helpful guide for all users. A second edition of the *NCBI Handbook* was made available in FY2014 with new chapters on MedGen and NCBI's eukaryotic and prokaryotic genome annotation pipelines.

The *NCBI Insights* Blog continues to be populated with helpful information for users, such as

sequence and genome build updates, instructional articles related to use of NCBI databases and tools, and general science news. The NCBI News site provides quick updates and news on all NCBI services and resources.

NCBI's social media presence continues to increase as usage of NCBI's Facebook, Twitter, and YouTube sites increases. NCBI added many tutorial videos to its YouTube site this year including:

- Submitting Manuscripts on NIHMS
- Sequence Viewer: PDF Rendering Feature
- BLAST: Create Custom Databases for Web BLAST
- The Variation Viewer
- Downloading FASTA Sequences in Sequence Viewer
- BLAST in the Cloud (Webinar)
- dbGaP: Complete a Study Registration
- dbGaP: Apply for Controlled Access Data
- dbGaP: Renew Authorized Access
- dbGaP: Close out a Controlled Access Project

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.

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.

In partnership with the National Library of Medicine Training Center, NCBI offered *A Librarian's Guide to NCBI* in April 2014 on the NIH campus. This course prepares health science librarians for supporting and training their patrons about NCBI databases and tools. This increasingly popular course provides two stages of instruction. A pre-course is available online to provide information on the fundamentals of bioinformatics and searching. The second part is a five-day course held on the NIH campus. The course is a combination of instruction, demonstration, discussions, and hands-on exercises.

Webinars have become increasingly popular, and are offered monthly, free of charge, by NCBI staff. Webinar topics in FY2014 included: Variation Viewer, BLAST, OSIRIS, E-Utilities, and variation clinical resources. 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. Analyzing user queries allows NCBI to improve and add sensors that detect particular query patterns, and respond accordingly.

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

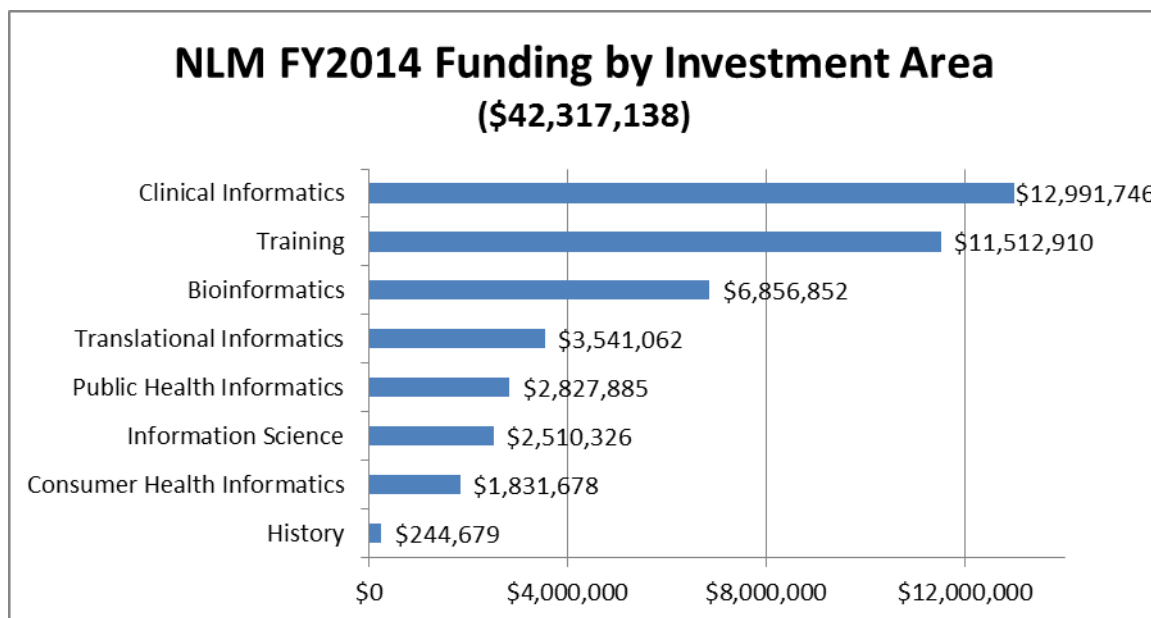
makes new and/or continuing awards in these five grant categories: Research Projects, Resources, Career Development, Research Career Training, and Small Business Research & Development.

Overview of FY2014

NLM's EP FY2014 base budget for grant awards was \$42,317,138, an increase of 2.7 percent over the FY2013 sequestration level but still nearly \$2 million below the FY2012 level. As has been the case in recent years, the fiscal year began under a Continuing Resolution. NIH Institutes and Centers cannot make new awards or launch new initiatives until an appropriations bill or other annual funding bill has been passed. One hundred twenty-nine new and continuing awards were made with NLM's appropriated funds in FY2014.

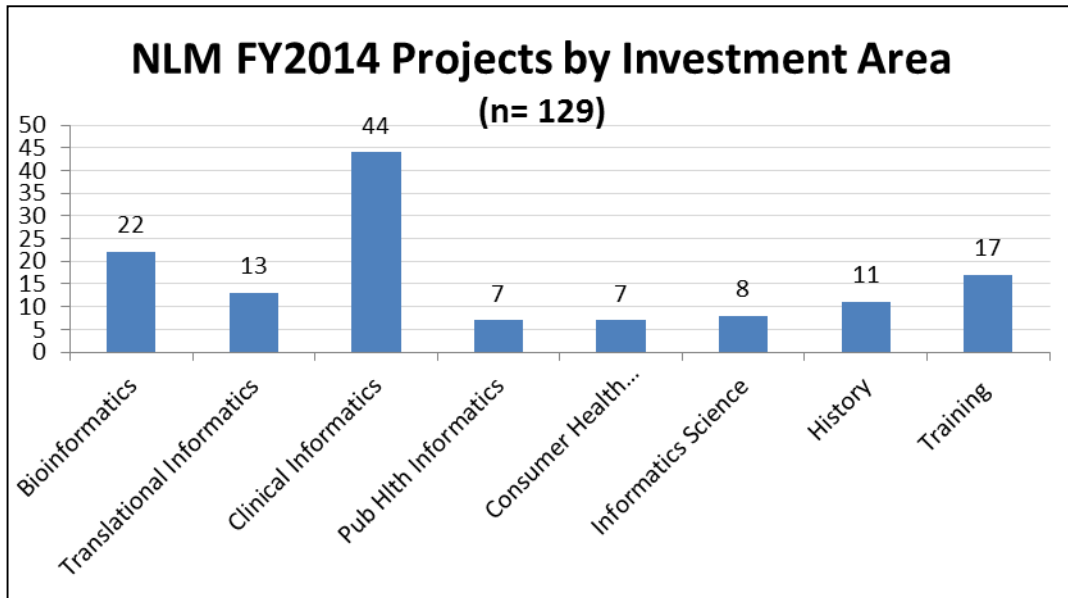
Figure 1 shows the distribution of NLM's FY2014 grant expenditures by grant types and Figure 2 shows the distribution by investment areas defined by program staff. Appendix 4 provides a detailed breakdown by mechanism and activity of NLM's FY2014 grant budget expenditures.

Figure 1. NLM FY2014 Funding by Investment Area²⁶



²⁶ Data for FY2014 NLM funding by Investment Area were collected from Query, View and Report (QVR), https://apps.era.nih.gov/qvr/web/home_main.cfm, at NIH on February 3, 2015.

Figure 2. NLM FY2014 Projects by Investment Area ²⁷



²⁷ Data for FY2014 NLM projects by Investment Area were collected from Query, View and Report (QVR), https://apps.era.nih.gov/qvr/web/home_main.cfm, at NIH on February 3, 2015.

Highlights

Budget reductions and an increase in the number of research grant applications submitted to NIH created a drop in success rates for all NIH Institutes and Centers in the past three years. Success rates are computed by dividing the number of awards by the number of applications reviewed in a fiscal year. Research grant success rates show the most effect. In 2013, NLM’s grant staff took action to improve research grant success rates by reallocating funds from other grant programs. **Table 6** shows that success rates improved for all of NLM’s core grant programs in FY2014 as compared to FY2012, FY2013. Success rates for all NIH Institutes and Centers are posted at http://report.nih.gov/success_rates/index.aspx

Table 6. Success Rates—Core NLM Grant Programs, FY2012-2014

	Activity Code	FY2012	FY2013	FY2014
Research	R01	14%	16%	21%
	R21	11%	3%	16%
Career	K22	50%	50%	43%
	K99	26%	25%	25%
	K01	N/A	N/A	60%
Resource	G13	8%	7%	13%

Eleven new awards were made in NLM’s 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, the first year of this program. In October 2014, the Association of Academic Health Sciences Libraries offered a three-part continuing education Webinar series for its members featuring 2012 awardees of this program talking about their experiences. For the 2014 program, 12 other NIH Institutes joined NLM’s announcement, thereby making their grantees eligible. Of the eleven awards made, six received partial funding from the parent Institute. Four of the eleven awards made in FY2014 went to institutions that also received 2012 awards (but for different projects).

Although NLM did not announce any new grant programs of its own in FY2014, participation in NIH-wide programs yielded new awards for high-priority informatics topics. As of FY2014, a new NIH Director’s Pioneer Award was assigned to NLM, Sean Wu’s “Enabling Technologies for Human-Machine Hybrid Tissues.” The NIH Common Fund provided year first funding; future year awards will be jointly funded by NLM and the NIH Common Fund.

NLM funded two new awards from the NIH Director’s Office FY2014 RFA “Empirical Research on Ethical Issues to Central IRBs” and “Consent for Research Using Clinical Records and Data.” Both awards (Beskow and Lee) deal with patient preferences regarding consent for the use of clinical data and samples. The first year of one of these grants was fully funded by the NIH Office of the Director.

NIH’s Big Data to Knowledge (BD2K) initiative issued a number of funding announcements that led to FY2014 awards that benefit NLM’s existing grantees. Eleven Centers of Excellence for Big Data Computing were awarded, four of them located at institutions with NLM Informatics training programs and research grant recipients. NLM program staff serve as lead science officer or science officer on five of these BD2K Center awards. Two BD2K supplemental awards went to NLM research project grantees (Hripcsak and Cohen) for pilot projects relating to the Data Discovery Index Coordination Consortium (DDICC). In the area of BD2K Training and Career Development, a Funding Opportunity Announcement was issued for supplements to NLM’s existing training programs to support new training tracks in data science. Awards will be made in FY2015.

Outcomes and Impact of NLM Research Grants

According to Scopus, NLM grantees published 233 articles in FY2014 acknowledging NLM grant support. These articles were published in 159 distinct journals. **Table 7** shows the top ten journals in which NLM grantees published in 2014. Citations are a measure used at NIH to assess the near and longer term impact of grantee research. Articles published in 2014 were cited 258 times. Eighty-three percent of the 233 FY2014 articles have been cited once; seventeen percent of the FY2014 articles have been cited more than once.

Table 7. Top 10 Journals Where NLM Grantees Published in FY2014

Publication Source	Articles Published
Journal of the American Medical Informatics Association	18
Bioinformatics	9
BMC	5
Pacific Symposium on Biocomputing	5
Proceedings of the National Academy of Sciences of the United States of America	4
PLoS Neglected Tropical Diseases	3
Pharmacological Research	3
Journal of Community Health	3
Human Genetics	3
Medical Image Computing and Computer Assisted Intervention	3

The three most highly cited articles published in 2014 that acknowledge NLM grant support are:

- Genetics of Rheumatoid Arthritis Contributes to Biology and Drug Discovery. Nature 2014, Vol. 506, no. 7488. More than 100 authors including NLM Grantee Denny, JC. Cited 21 times in 2014.
- Quantifying the Local Resolution of Cryo-EM Density Maps. Nature Methods 2014, Vol. 11, no. 1. Three authors including NLM Grantee Tagare, HD. Cited 14 times in 2014.
- Genetic Data and Electronic Health Records: A Discussion of Ethical, Logistical and Technological Considerations. Journal of the American Medical Informatics Association 2014, Vol. 21, no. 1. Four authors including NLM Grantee Mendonca, EA. Cited 4 times in 2014.

The NLM active grant portfolio

At the end of FY2014, NLM’s active grant portfolio totaled 150 grants worth \$47.066 million. The NIH RePorter provides a graphic visualization of the active grants in **Figure 3**, showing distribution across topic areas that is not parallel. For example, “Gene” is not a significant topic in the active grants visual, but it stands out in the journal graphic.

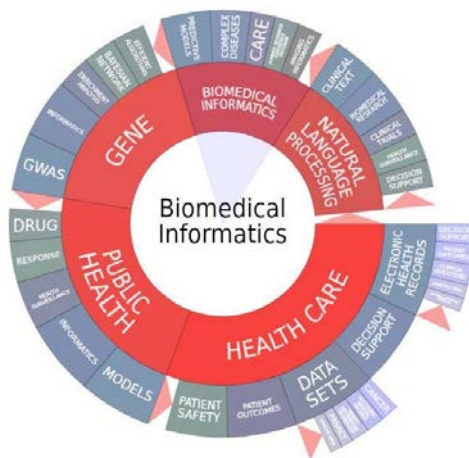
Figures 3 and 4 below is a Graphical View of NLM Active Grants by Topic Area and Journal Title²⁸

Figure 3. FY2014 150 Active NLM Grants by Topic



²⁸ Active grant data was collected from Scopus, <http://www.scopus.com/list/form/overview.url>, at NIH on February 6, 2015.

Figure 4. FY2014 Projects by Topic Area



The three most highly cited articles from NLM’s active portfolio published within the past 5 years:

- An Integrated Map of Genetic Variation from 1,092 Human Genomes. *Nature* 2012, Vol. 491, no. 7422. More than 100 authors, including NLM fellow Maples, BK from Stanford’s NLM Informatics Training Program. (Several staff members of NCBI are also major contributors). Cited 613 times.
- Emerging fungal threats to animal, plant and ecosystem health. *Nature* 2012, Vol. 484, no. 7393. Seven authors including NLM grantee Brownstein, JS. Cited 126 times
- Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease. *Nature Reviews Genetics* 2010, Vol. 11, no. 6. Seven authors including NLM grantee Moore, JH. Cited 112 times

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.

- 15 new R01 awards (12 in FY2013); 43 continuing R01 awards

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.

- 6 new R21 awards (1 in FY2013); 1 continuing R21 award

Conference Grants

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

- 2 new R13 awards (0 in FY2013); 2 continuing R13 awards

Small Business Innovation Research and Small Business Technology Transfer (SBIR/STTR) 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 2.8 percent for SBIR and 0.4 percent for STTR. NLM's SBIR/STTR interests for FY2014 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 FY2014, EP received applications through the NIH Omnibus Solicitation, and through a trans-NIH BISTI FOA.

- One R43, 0 R44 new SBIR awards (1 R43, 1 R44 in FY2013); one continuing SBIR award

NIH Academic Research Enhancement Awards (AREA) Grants

The purpose of the NIH R15 Academic Research Enhancement Award (AREA) program is to stimulate research in educational institutions that provide baccalaureate or advanced degrees for a significant number of the Nation's research scientists, but that have not been major recipients of NIH support.

- One new R15 award (0 in FY2013)

NIH High Risk-High Reward Grant Programs

These programs are operated by the Common Fund on behalf of all NIH Institutes and Centers. Typically, a certain number are fully or partially funded by the Common Fund in the first year, with remaining years of support coming from the Institute to which the grant is assigned.

- In FY2014, NLM became home to a five-year NIH Director's Pioneer award given to Dr. Sean Wu at Stanford University for "Enabling Technologies for Human-Machine Hybrid Tissues." The award is jointly funded by the Common Fund and NLM.
- In FY2014, NLM became home to a five-year NIH Director's New Innovator Award given to Dr. Bree Aldridge at Tufts University for "Quantitative Design of Multi-drug Regiments for Tuberculosis." The award is fully funded by the NIH Common Fund.

Resource Grants

Resource Grants use appropriated funds to support dissemination and management of health-related information. These grants are not research grants. 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)

This program solicits applications every other year. The purpose of the program is to bring useful, usable health information to health disparity populations and their health care providers. The program strongly encourages applications that involve health sciences libraries and minority-serving institutions. Applications received from RFA LM 14-001 were reviewed by NLM in FY2014, for awards to be made in early FY2015.

- 0 new G08 information resource grants (3 in FY2013); 3 continuing G08 awards

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.

- 5 new G13 Scholarly Works awards (3 in FY2013); 6 continuing G13 awards

Administrative Supplements for Informationist Services in NIH-funded Research Projects

These two-year awards are issued as administrative supplements to existing research grants of other NIH Institutes and Centers. Informationists, who are often librarians, work as team members in the research team to assist with acquisition, synthesis, management, and use of research data.

- 11 new awards (0 in 2013); 0 continuing awards

Training and Career Awards

NLM remains the principal US source of support for research training in biomedical informatics. NLM provides both institutional training support and individual development transition support. In addition to the NLM-initiated programs listed below, in FY2014, NLM participated in several NIH Big Data to Knowledge (BD2K) informatics and data science training initiatives.

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

Five-year institutional training grants support pre-doctoral and post-doctoral research fellowships at university-based programs across the country.

NLM's fourteen university-based informatics training programs are located at the following US organizations:

1. University of California San Diego (San Diego, CA)
2. Stanford University (Stanford, CA)
3. University of Colorado Anschutz Medical Campus (Aurora, CO)
4. Yale University (New Haven, CT)
5. Harvard University (Medical School) (Boston, MA)
6. Columbia University Medical Center (New York, NY)
7. Ohio State University (Columbus, OH)
8. Oregon Health & Science University (Portland, OR)
9. University of Pittsburgh (Pittsburgh, PA)
10. Vanderbilt University (Nashville, TN)
11. Rice University (Houston, TX)
12. University of Utah (Salt Lake City, UT)
13. University of Washington (Seattle, WA)
14. University of Wisconsin Madison (Madison, WI)

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) supports pre- and post-doctoral trainees in dental informatics at three of NLM's training programs: University of Pittsburgh, Oregon Health & Science University and the University of Wisconsin-Madison.

In FY2014, NLM supported 117 pre-doctoral and 93 postdoctoral fellowship slots at its 14 university-based training programs, of which 2 pre-doctoral and 7 postdoctoral trainees are in dental informatics and are supported by the National Institute of Dental and Craniofacial Research (NIDCR).

The annual NLM Informatics Training Conference was held at the University of Pittsburgh, in Pittsburgh, PA, on June 17-18, 2014. Approximately 248 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; and guests. Research projects were presented in plenary and parallel sessions by 31 informatics trainees. There were 13 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 29 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: Patricia Yao, Oregon Health & Science University, "Texting for Sexual Health to American Indian/Alaska Native Teens and Young Adults."
- Best Poster Day 1: Corey Hryc, Rice University/Baylor, "Validation of Near-Atomic Resolution Molecules to Identify Structures for Drug Design".

- Best Poster Day 2: Sivan Kinberg, Columbia University, "Steroid-Induced Diabetes Mellitus in Patients with Inflammatory Bowel Disease."
- Best Presentation Day 1: Shomir Chaudhuri, University of Washington, "Older Adults' Attitudes Toward Fall Detection Devices."
- Best Presentation Day 2: Rafael Rosengarten, Rice University/Baylor, "Predicting Gene Regulatory Interactions by Intermediate Data Fusion."

Details of the agenda and a photo gallery are available at <http://www.nlm.nih.gov/ep/TrainConf2014/TrainingConf2014.html>.



Attendees at the NLM Informatics Training Conference, June 2014 in Pittsburgh, PA.

Ruth L. Kirschstein NRSA Individual Pre-doctoral Fellowships (F30/31) F3

On recommendation of the Advisory Committee of the NIH Director, all NIH Institutes and Centers began offering the NRSA Individual Pre-doctoral Fellowship, as an alternative to traditional training offered by a university-based training program. These fellowships allow a would-be trainee to propose a curriculum, mentorship plan and research that will lead to a PhD at any US university. NLM has received its first F31 application, to be considered for funding in FY2015.

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 independent 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.

- 2 new K99 awards and 2 new R00 awards were made in FY2014 (2 new K99 awards and 3 new R00 awards in FY2013). 8 continuing R00 awards.

K01 Career Development Award in Biomedical Informatics

In FY2013, NLM issued a new career development FOA (K01) to replace our *NLM Independent Career Development Award for Biomedical Informatics (K22)*, which expired and could not be reissued for administrative reasons. The new FOA provides support for promising junior investigators as they launch their research careers in biomedical informatics research.

- Three new K01 awards

K22 NLM Independent Career Development Award for Biomedical Informatics

Between FY2010 and 2013, NLM offered the K22 career transition award. As noted above, it was replaced in FY2014 at the request of NIH.

- Three K22 awards made before the program closed; 4 continuing awards

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). Five-year awards were made in FY2011 to University of California Irvine, Vanderbilt University and Oregon Health Sciences University. In FY2014, these awards continued to support eight summer trainee slots at each program site.

Pan-NIH Projects and Interagency Collaborations

National Centers for Biomedical Computing (NCBC)

The National Centers for Biomedical Computing (NCBC) were cooperative agreement awards originally funded under the NIH Roadmap, but transitioned to Institute funds across the final five-year period. In 2013, NLM's NCBC center, "Informatics Integrating the Bench and Bedside (i2b2)," based at Harvard University's Brigham and Women's Hospital, ended its funding cycle. NLM continues to provide co-funding for the NHLBI NCBE "Integrating Data for Analysis, Anonymization and SHaring (iDASH)" at University of California San Diego, the final remaining NCBC, whose funding cycle ends in FY2015.

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 2. In FY2014 an AREA grant was awarded to Keyuan Jiang in the amount of \$359,922 for her project detecting drug effects using social media. 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

In FY2014, NLM provided co-funding support for eleven informationist supplements to grants of other NIH Institutes, an STTR award administered by NIMH, and the iDASH NCBC administered by NHLBI. The total cost of these co-funds from NLM to other Institutes is: \$167,586.

In addition, NLM received funding from other organizations in support of grants it administers. The National Institute of Dental and Craniofacial research continues to provide support for nine dental informatics trainees at NLM training programs. Two NLM grantees received BD2K supplement funds for pilots relevant to the Data Discovery Index Coordination Consortium (DDICC), NIGMS provided funding for one NLM grant through its Institutional Development Award (IDeA) grant program, which supports research at an institution in a state that does not receive significant funding from NIH grants. The NIH OD provided year-one funding for one research grant related to consent policy in the use of personal health information. The NIH Common Fund provided year one funding for the DP1 Pioneer award to Sean Wu and also continuation funds for the DP1 Pioneer award to Ram Samudrala. Six NIH Institutes provided partial funding for Informationist Supplements. These NIH co-funding initiatives from other Institutes to NLM add up to a total of \$3,405,430.

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 \$100,000 in FY2014. The PDB, which is administered by the National Science Foundation (NSF), is the single world-wide repository for the processing and distribution of 3-D biological macromolecular structure data.

NSF and seven NIH Institutes, including NLM, issued a joint funding solicitation for research projects entitled “Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA). Hundreds of applications were reviewed by NSF study sections; reviewers used both NSF and NIH criteria for applications with a health or biomedical research focus. NLM chose two projects from this initiative to transfer to NIH for funding consideration as a research project grant in FY2014; one was funded.

Extramural Programs Web site

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

In FY2014, there were 82,731 visitors to the EP Web site, a 15 percent increase over the previous year. The EP Web site continued to have a substantial interest from international visitors who represented almost 20 percent of total visits. The most frequently viewed pages were those describing the grant programs, EP homepage, information pages such as the NLM FY2014 Funding Plan and Grant Deadlines, Frequently Asked Questions, and a page of links to NIH-produced tutorials for preparing grant applications. Although a new program, the EP’s Big Data to Knowledge (BD2K) page ranked 10th in most frequently browsed Web pages.

Presentations by Grantees

Program staff organized a series of Informatics Lectures by NLM grantees. Dr. Timothy Cardozo presented on “Matching Complex Biomarkers to Drugs Using HistoReceptomic Signatures,” on November 13, 2013. Dr. Chunhua Weng presented on “Bridging the Semantic Gap between Research Eligibility Criteria and Clinical Data,” on March 5, 2014. Dr. Guergana Savova presented on “Temporal Relation Discovery from the Clinical Narrative,” on June 4, 2014. These lectures were held at Natcher Conference Center, Balcony A, and were well attended by NLM and NIH staff.

Scientific Review

Overall, 209 applications were reviewed for which NLM was the primary assignment. Of those, 174 were reviewed by NLM. The remaining 35 applications were reviewed by CSR, including SBIR/STTR grant applications. Of the applications reviewed by NLM, 64 percent were in one of the three research grant mechanisms (72 R01s, 36 R21s, and 3 R13s). Career Development Awards (K99s, K22s) represented 11 percent (19 applications), Scholarly Works (G13s) at 25 percent (44 applications) of the total applications reviewed.

NLM’s standing review group, the Biomedical Library and Informatics Review Committee (BLIRC), evaluates grant applications assigned to NLM for possible funding. BLIRC met three times in FY2014 and reviewed 104 applications. The Committee (**Appendix 10**) reviews most of the applications for biomedical informatics and bioinformatics research, career support, and fellowships.

Four Special Emphasis Panels were held during FY2014. These panels are convened on a one-time basis to review applications for which the BLIRC lacks appropriate expertise, such as Scholarly Works grant applications, or when a direct conflict of interest exists between the application and a member of the BLIRC. Overall, NLM’s SEPs reviewed a total of 70 applications during FY2014, compared to 96 in FY2013.

Concurrence with the results of initial review, called second-level review, is performed by the Board of Regents (BOR). The BOR Extramural Programs Subcommittee conducts early concurrence reviews electronically on the most fundable research grants and on special initiatives such as the Training Grant RFA. In FY2014, the subcommittee held three early concurrence panels; 30 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 1,079 NLM grant applications were included in the en bloc votes (209 primary and 870 secondary).

In an effort to cut review related administrative costs, EP continued the use of teleconferencing for large SEP meetings. The teleconferencing format was used in a two-day long session to conduct the review of 40 Scholarly Works applications. The cost of review per application was approximately an 8-fold savings when compared to using face-to-face meeting format.

Appendix 1: RFA/PA Actions in FY2014 - 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	NLM Informatics Conference Grants (R13)	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-14-339	NLM Grants for Scholarly Works in Biomedicine and Health (G13)	February 23, 2016
PAR-13-284	NLM Career Development Award in Biomedical Informatics (K01)	September 8, 2016
PA 15-083	NIH Pathway to Independence Award (K99/R00)	January 8, 2017

Appendix 2: RFA/PA Actions in FY2014 - Multi-Institute Active Announcements in which NLM Participates

Announcement	Title	Expiration
PA-12-150	Research Grant Supplement to Promote Reentry in Health-Related Research (Admin Supp)	September 30, 2015
PA-12-149	Research Grant Supplement to Promote Diversity in Health-Related Research (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-RM-13-006	NIH Pioneer Award Program (DP1)	October 15, 2015
RFA-RM-13-007	NIH Director's New Innovator Award Program (DP2)	October 17, 2015
RFA-HG-14-007	Mentored Career Development Award in Biomedical Big Data Science for Clinicians and Doctorally Prepared Scientist (K01)	April 2, 2015
RFA-HG-14-008	Courses for Skills Development in Biomedical Big Data Science (R25)	April 2, 2016
RFA-HG-14-006	Revisions to Add Biomedical Big Data Training to Active NLM Institutional Training Grants in Biomedical Informatics (T15)	July 26, 2016
RFA-HG-14-004	Predocctoral Training in Biomedical Big Data Science (T32)	July 28, 2015
PAR-14-071/072	Small Business Innovations Research Grants (SBIR R43/R44) (STTR R41/R42)	May 8, 2015

Extramural Programs

Announcement	Title	Expiration
PA-12-196	Innovative Health Information Technology for Broad Adoption by Healthcare Systems and Consumers (SBIR)(R44)	January 8, 2015
PA-14-154/157	Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science (R43/R44) (R41/R42)	May 8, 2017
PA-14-147/148/150	Ruth L. Kirschstein NRSA Individual Predoctoral Fellowships (F30/F31)	January 8, 2017

Appendix 3: NLM New FY2014 Awarded Grants

NLM NEW FY2014 AWARDS

Research Projects (R01)

Beskow, Laura M
Ethical Approaches to Research Use of Clinical Records and Data
1 R01 LM012178-01
Duke University

Boyce, Richard
Addressing Gaps in Clinically Useful Evidence On Drug-Drug Interactions
1 R01 LM011838-01
University of Pittsburgh at Pittsburgh

Burd, Randall S
Automatic Workflow Capture & Analysis for Improving Trauma Resuscitation Outcomes
1 R01 LM011834-01A1
Children's Research Institute

Culley, Joan M
Validating Triage for Chemical Mass Casualty Incidents - A First Step
1 R01 LM011648-01A1
University of South Carolina at Columbia

Denny, Joshua C
From GWAS to PheWAS: Scanning the EMR Phenome for Gene-disease Associations
2 R01 LM010685-04
Vanderbilt University

Gennari, John H
Physiological Knowledge Integration and Recombinant Modeling Via Accelerated Sema
1 R01 LM011969-01
University of Washington

Hunter, Lawrence E
Developing and applying information extraction resources and technology to create
2 R01 LM008111-09A1
University of Colorado Denver

Jiang, Xia
A New Generation Clinical Decision Support System
1 R01 LM011663-01A1
University of Pittsburgh at Pittsburgh

Lee, Sandra S
Beyond consent: patient preferences for governance of use of clinical data and samples
1 R01 LM012180-01
Stanford University

Li, Lang
Evidence-based Drug-Interaction Discovery: In-Vivo, In-Vitro and Clinical
1 R01 LM011945-01
Indiana Univ-Purdue Univ at Indianapolis

Liu, Hongfang
Semi-structured Information Retrieval in Clinical Text for Cohort Identification
1 R01 LM011934-01A1
Mayo Clinic Rochester

Malin, Bradley A
Automated Detection of Anomalous Accesses to Electronic Health Records
2 R01 LM010207-05A1
Vanderbilt University

Sarkar, Indra N
In silico Identification of Phyto-Therapies
1 R01 LM011963-01
University of Vermont & St Agric College

Tao, Cui
Patient Medical History Representation, Extraction, and Inference from EHR Data
1 R01 LM011829-01A1
University of Texas Hlth Sci Ctr Houston

Wright, Adam
Improving Clinical Decision Support Reliability Using Anomaly Detection Methods
1 R01 LM011966-01
Brigham and Women's Hospital

Xie, Lei
Drug Discovery by Integrating Chemical Genomics and Structural Systems Biology
1 R01 LM011986-01
Hunter College

Xu, Hua
Interactive Machine Learning Methods for Clinical Natural Language Processing
2 R01 LM010681-05
University of Texas Hlth Sci Ctr Houston

Xu, Xiaoyin
A Novel Informatics Approach to Understanding Complex
Muscle Fiber Phenotypes
1 R01 LM011415-01A1
Brigham and Women's Hospital

Exploratory/Developmental Research (R21)

Arnold, Corey W
A Topic Model and Visualization for Automatic
Summarization of Patient Records
1 R21 LM011937-01A1
University of California Los Angeles

Gerek, Zeynep N
Computational diagnosis of non-synonymous variations
using structural dynamics
1 R21 LM011941-01
Temple Univ of the Commonwealth

Grando, Maria A
Patient-centered decision support based on device
evidence (I DECIDE)
1 R21 LM011667-01A1
Arizona State University-Tempe Campus

Jiang, Xiaoqian
Protect Privacy of Healthcare Data in the Cloud
1 R21 LM012060-01
University of California San Diego

Kalet, Ira J
Meta-modeling methods for biological processes
1 R21 LM012075-01
University of Washington

Wong, Carolyn
Visualizing Highway Pollution: A Study of Inter-
generational Health Communication
1 R21 LM011919-01A1
University of Massachusetts Boston

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

Boullerne, Anne I
Glial Cell Biology: A Historical Perspective
1 G13 LM011465-01A1
University of Illinois at Chicago

Kim, Daniel
New Horizons in Modeling and Simulation for Social
Epidemiology and Public Health
1 G13 LM012056-01
Northeastern University

Muller, Keith E
Power and Sample Size for Multilevel and Longitudinal
Designs in Health Research
1 G13 LM011879-01A1
University of Florida

Pollock, Ludmila
Human Genome Project Documentary History: An
Annotated Scholarly Guide to the HGP
1 G13 LM011890-01
Cold Spring Harbor Laboratory

Pomata, Gianna
At the Origins of the Case History: Individualized
Medicine in Pre-Modern Europe
1 G13 LM011447-01A1
Johns Hopkins University

NIH Director's Pioneer Award (NDPA) (DP1)

Wu, Sean M
Enabling Technologies for Human-Machine Hybrid
Tissues
1 DP1 LM012179-01
Stanford University

**Career Development Award in Biomedical Informatics
(K01)**

Hoen, Anne G
Bioinformatics strategies for early life microbiomics
1 K01 LM011985-01
Dartmouth College

Huh-Yoo, Jina
InfoMediator: Weaving Clinical Expertise in Online
Health Communities
1 K01 LM011980-01
Michigan State University

Soulakis, Nicholas D
A Novel Graph Processing Architecture to Ascertain &
Monitor Care Coordination
1 K01 LM011973-01
Northwestern University at Chicago

Career Transition Award (K22)

Beck, Andrew
Informatics Methods and Models For Computational
Pathology
1 K22 LM011931-01
Beth Israel Deaconess Medical Center

Hebbring, Scott J
Development and Application of Phenome-wide Scan of Heritability (PheSH)
1 K22 LM011938-01
Marshfield Clinic Research Foundation

Starkey, Jonathan M
Predictive Model of Chronic Kidney Disease in a Hispanic Population
1 K22 LM011869-01A1
University of Texas Medical Br Galveston

NIH Pathway to Independence Award (K99)

Lu, Songjian
Developing Graph Models and Efficient Algorithms for the Study of Cancer Disease
1 K99 LM011673-01A1
University of Pittsburgh at Pittsburgh

Waghlikar, Kavishwar B
A Framework to Enhance Decision Support by Invoking NLP: Methods and Applications
1 K99 LM011575-01A1
Mayo Clinic Rochester

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

Klassen, Daniel L
Memory Matters: A Mobile Aid to Stimulate Reminiscing in Persons with Memory Loss
1 R43 LM012085-01
Moai Technologies, LLC

NLM Informatics Conference (R13)

Dunker, Alan K
Pacific Symposium on Biocomputing
2 R13 LM006766-18
Indiana Univ-Purdue Univ at Indianapolis

Weiner, Jonathan P
Symposium on Population Health Informatics
1 R13 LM011955-01
Johns Hopkins University

Academic Research Enhancement Award (AREA) (R15)

Jiang, Keyuan
Detection of Potential Drug Effect Signals from Twitter Data
1 R15 LM011999-01

NIH High Priority, Short-Term Project Award (R56)

Shatkay, Hagit
Utilizing Imaged-based Features in Biomedical Literature Classification
1 R56 LM011354-01A1
University of Delaware

Appendix 4: FY2014 Grant Budget by NIH Mechanism

FY2014 Grant Budget by NIH mechanism	No.	Total
Research Project Grants (R01, R21, R00, R15)	78	\$27,150,574
SBIR/STTR (R41, R42, R43, R44)	2	\$799,326
Other Research - Research Careers (K99, K22, K01)	12	\$1,776,727
Other Research - Other (G08, G13, R13, R25)	21	\$927,601
Training - Institutional (T15)	14	\$11,512,910
R&D Contracts (Y03)	1	\$100,000
National Centers for Biomedical Computing Award (U54)	-	\$50,000
EP Grant budget excluding TAPS and Operations	129	\$42,317,138

Appendix 5: FY2014 Grant Budget by NIH Activity Code

FY2014 Grant Budget by NIH activity code	No.	Total
DP1: Pioneer Award	1	\$ - *
G08: Information Resource Grants to Reduce Health Disparities	3	\$203,104
G13: Scholarly Works in Biomedicine and Health	11	\$488,873
K99: Pathway to Independence	2	\$275,873
K22: NLM Independent Career Development Award for Biomedical Informatics	7	\$1,051,135
U54: National Center for Biomedical Computing	-	\$50,000
K01: Career Development Award in Biomedical Informatics	3	\$449,719
R00: Pathway to Independence	9	\$2,018,535
R01: Research Project Grants	61	\$23,247,003
R13: Conference Grants	4	\$52,499
R15: Academic Research Enhancement Award (AREA)	1	\$359,922
R21: Exploratory/Developmental Grants	7	\$1,525,114
R25: Education Projects	3	\$183,125
R41: Small Business Technology Transfer (STTR)	-	\$ -
R42: Small Business Technology Transfer (STTR)	-	\$98,222
R43: Small Business Innovation Research (SBIR)	1	\$235,904
R44: Small Business Innovation Research (SBIR)	1	\$465,200
T15: University Biomedical Informatics Research Training Programs	14	\$11,512,910
Y03: Inter-Agency Agreement	1	\$100,000
EP budget excluding TAPS and Operations	129	\$42,317,138

*This NIH Director's Pioneer Award (DP1) is a co-funding collaboration between the NIH and the NLM. The initial year is supported by the NIH—Common Fund—and subsequent years are supported by both the NIH and the NLM.

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.

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

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

Controlled Medical Vocabularies

Unified Medical Language System® (UMLS) Project

We provided updated medical terminology content in two scheduled releases of the Unified Medical Language System (UMLS)—an NLM information resource that integrates and distributes key medical terminology, classification and coding standards. The 2013AB release included four new sources; 2.9 million concepts, an annual increase of 1.53 percent over 2012AB; and 9.3 million concept names, an annual increase of 2.62 percent over 2012AB. The 2014AA release included one new terminology, "SNOMED CT Veterinary Extension" (SNOMEDCT_VET); 2.97 million concepts, an increase of 1.46 percent over 2013AB release; and 11.63 million concept names, an increase of 2.04 percent over 2013AB.

We also leveraged the Terminology Services inversion framework to enable standardization and code reuse across different ontologies, and migrated nine ontologies to use this framework. Inversion is the process that normalizes codes, relationships and other pertinent information from medical terminology sources before they are incorporated into the UMLS.

We also automated our quality assurance (QA) checks for the inversion process that normalizes codes, relationships and other pertinent information from medical terminology sources before they are ingested into the Unified Medical Language System (UMLS). By automating the QA functionality, we stand to prevent several weeks of potential delays in the release of future versions of the UMLS by uncovering errors early in the production cycle.

We added new features to MEME (Metathesaurus Enhancement and Maintenance Environment), the editing environment for the UMLS. MEME is comprised of a set of editing, quality assurance, maintenance, and administrative tools that manage and support all of the processes required to edit medical ontologies to a state where it is ready for production. New features included changes and additions to the categories into which a concept (editing unit) can be placed. The new categories will assist editors so they can more easily view and work on a more functional level. We also made reporting enhancements.

Value Set Authority Center (VSAC) Project

We made significant improvements to the Value Set Authority Center (VSAC)—an online portal that serves as the source of truth for medical terminology used in the Secretary's initiative for clinical quality measurement. Some of our notable accomplishments included implementing a new authoring environment; developing a value set dashboard to improve the value set curation process; establishing a shared environment to enhance the value set review process; making enhancements to the code context information in order to improve the productivity of value set authoring; improving code validation to maintain the integrity of value sets; adding new value sets to VSAC to support Meaningful Use Stage 2 (MU2) Certification annual updates; and adding several new code systems to enhance the value set authoring.

We created new functionality that gives Centers for Medicare and Medicaid Services (CMS) sponsored authors the ability to create, edit, and maintain the sets of medical codes that are required for MU2 Certification. This new capability will expedite authoring by several months every year. In addition, four updates to the terminology content were release this fiscal year, in accordance with CMS regulation.

In addition, we provided the VSAC authors with the ability to find desired codes by selecting root concepts

within the hierarchy of a medical terminology. We implemented this tree browsing capability for the following code systems: Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), International Classification of Diseases book 9, Clinical Modification (ICD-9-CM), International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), International Classification of Diseases, Tenth Revision, Procedure Coding System (ICD-10-PCS), and Source of Payment (SOP). This tree navigation functionality allows authors who are less proficient with medical code systems to be able to develop value sets for the MU2 certification program.

Moreover, we added functionality to VSAC that allows authors to remap inactive and obsolete medical codes into codes that are active. This functionality prevents published value sets from containing stale codes and substantially reduces the efforts of authors in maintaining up-to-date codes in their corresponding value sets.

We created a value set release dashboard to give CMS-sponsored reviewers the capability to load, select and validate value sets prior to each release specified in CMS regulation. The dashboard automates the identification of problems within value sets and provides suggestions for improving them. This functionality not only improves the quality of value sets, but also significantly reduces the time required for reviewing value sets prior to publication.

We extended VSAC by adding support for seven new HL7 code systems, thus providing authors the ability to select codes from the newly added code systems when creating value sets for the MU2 program.

International Health Services Terminology Standards Development Organization (IHTSDO) Project

We implemented a SNOMED-CT International Request Submission System (SIRS) for the International Health Terminology Standards Development Organization (IHTSDO) to track new requests submitted by the user community for addition of medical concepts, descriptions and relationships to the International edition of SNOMED-CT. Prior to this implementation, IHTSDO users had no way to track the status of their requests. With SIRS, users are now able to track the progress of the request through each stage of review all the way through final approval. There have been about 16,538 requests created in SIRS for this year, and SIRS had 18,774 hits in September from 663 visitors.

The main improvement for the US SNOMED CT Content Request Submission (USCRS) system this year included the new functionality of allowing interested parties to deploy customized versions of the USCRS application for various uses. The SIRS application is one example.

RxNorm Project

We re-engineered the platforms for the RxNorm Editing System, which allows editors the functionality to manage content in NLM's flagship drug terminology product. This improved the backend system processing time by 93 percent, from 15 seconds to one second. This has also led to improved system reliability.

We released five new versions of the RxNorm Editing System, which provides the editors of RxNorm the tools needed to perform their tasks more quickly, with more power to manipulate the data and create and maintain new data types. In order to ensure continuity of operations for the RxNorm Editing system and release process, we enabled Active/Passive failover support at the offsite disaster recovery location. By load balancing across sites, we are better able to maintain uninterrupted service for our customers, and are able to conduct either planned or unplanned maintenance without service outages.

Additionally, we analyzed and refined the algorithms used to mark drug strings and NDCs as obsolete by leveraging additional data from our source providers. This resulted in a more accurate view of drugs that are on or off the market.

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 implemented a new Web-based editing system that provides controlled data entry and immediate field validation to prevent basic errors. The MeSH Translation Maintenance System (MTMS) and the Global Citation Management Systems (GCMS) were also updated to take advantage of the new MeSH editing interface.

We improved performance of Year End Processing by reducing the time to process from seven hours to two hours, a 72 percent decrease in processing time, which increased the performance of all MeSH processes, including those related to MTMS and Descriptor and Chemical cutover.

We provided twelve updates to the MEDIT Editing System after initial release. These updates provided additional features and user interface changes in order to facilitate faster editing of each record and greater functionality for the editor to manipulate the data.

Additionally, we coordinated with Spanish/Portuguese MeSH users to load their data into MTMS for creation of their vocabulary solely inside of the MTMS system going forward. This will remove the burden of needing to load the Spanish/Portuguese data manually each year.

Common Data Elements (CDE)

Use of Common Data Elements (CDEs) in clinical research studies and electronic health records has the potential to enable more efficient research studies, efficient aggregation and reuse of clinical research data, and more rapid implementation of research findings in clinical practice. With funds initially provided to NLM and the Office of the National Coordinator for Health Information Technology (ONC) from the Patient Centered Outcomes Research Trust Fund, OCCS, we worked with Library Operations, the Trans-NIH Biomedical Informatics Coordinating Committee, as well as other HHS agencies (CMS, ASPE, AHRQ, etc.) to implement a CDE Repository. This scalable platform allows users to browse, search, author, and curate or perform comparative analysis of NIH- and HHS-funded CDE and form content, including NCI's caDSR, NINDS common data elements, PhenX data elements, PROMIS, AHRQ-common format for reporting patient safety issues, etc. NLM will leverage its biomedical terminologies repertoire to further refine and add capabilities and content to this platform to enhance the users' experience in the reuse, authorship and/or review of high-quality data elements and forms in accordance with their needs and usage.

Terminology Service (TS)

Terminology Service (TS) provides access to the data related to code systems (vocabularies). It hosts multiple versions of various code systems that are needed by UMLS and VSAC that in turn provide value-added service of the data. The TS provides the repository for all code systems hosted by NLM and serves the code system related data to its downstream applications. The following are three main downstream consumptions of TS data:

- UMLS—providing access to concepts of synonyms from various code systems,
- VSAC—providing access to value sets containing code system codes for Meaningful Use,
- Common Terminology Services 2 (CTS2) Service—providing API access to code system related data by implementing the CTS2 specification.

This year we developed the framework to enable TS services, added several new code system data sets to enrich the terminology data store, and implemented the code system catalog service.

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 applied new technology to the MedlinePlus Input System which improved performance time by nine percent and reduced data size by 44 percent.

We enhanced the MedlinePlus topics search module by adding a new feature to search by release status. Additionally, we have completed the implementation for the one-column view of the MedlinePlus responsive site to deliver better digital services. The new responsive site will be replacing the MedlinePlus Mobile site.

We worked in conjunction with the Web Support team to set up a new Linux server with ColdFusion version 10. At the end of the testing phase, we were able to dramatically improve the MedlinePlus overnight build process from two and a half hours, down to 12 minutes, a 92 percent reduction in time. Additionally, we implemented application security logging in order to comply with NIH security policy standards and to provide a secure browsing experience for the public.

MedlinePlus Connect®

We expanded the existing Web application programming interface (API) for MedlinePlus Connect by adding support for the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code system. This initiative provides a mapping from ICD-10-CM codes to MedlinePlus Health Topics, helping electronic health record (EHR) providers achieve compliance for “meaningful use” of Health Information Technology.

The NLM automated the mapping of SNOMED CT codes to 1,430 MedlinePlus English and Spanish Health Topics. This initiative increased SNOMED CT vocabulary coverage in MedlinePlus Connect by 360 percent, from 13,000 to 60,000 terms, by leveraging mappings available for both the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) and ICD-10-CM. The mapping tool reportedly saved an estimated four months of staff time. It also made this SNOMED CT expansion effort possible, as without this tool, this effort might not have been attempted manually. Additionally, we re-engineered the API to support the Health Level Seven (HL7) Context-aware Knowledge Retrieval (Infobutton) standard.

We also created a MedlinePlus Connect mapping management system. This management system allows NLM's MedlinePlus team to more easily modify and monitor the data, such as import/export mapping. This automates the current cumbersome manual effort, increases the efficiency of the current work flow, and minimizes user input error.

DailyMed Project

DailyMed provides high quality information about marketed drugs. This information includes FDA-approved labels (package inserts). This Web site provides health information providers and the public with a standard, comprehensive, up-to-date, look-up and download resource of medication content and labeling as found in medication package inserts.

We redesigned the DailyMed Web site using the responsive Web design framework to be more mobile device friendly. DailyMed, like many of NLM's flagship Web applications, is beginning to see an increased percentage of mobile visitors. So, the new responsive design will ensure greater accessibility for mobile users of DailyMed.

We created 12 new Web APIs to DailyMed to provide more resources for users to access. We also implemented a safety alerts system to provide timely alerts on DailyMed.

NLM Digital Repository Project

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 added support for serial content. A new data model was designed to fully represent serials in a hierarchical fashion, including physical issues and logical entities such as volumes and parts. Software was developed to logically represent, index and display these entities. The new serials interface provides a consistent, unified approach to access, describe, search within, and view serials. We also deployed a full redundant architecture to the NCCS offsite data center, deployed nine new software releases with numerous functional and stylistic enhancements, and investigated enhancements to digitization hardware, software, workflow automation and video delivery.

As described in the Library Operations chapter, substantial new content was added to the NLM Digital Repository this year. In FY2014, NLM Digital Collections handled more than 5,600 Web service requests, 190,000 unique visitors and over 4.8 million page views, which is 12 percent more visitors and more than three times the number of page views than the previous fiscal year.

Digitization Scan Quality Assurance (QA) System

The Digitization Scan QA system is a Web-based quality assurance system for the NLM Digitization Program used to automate, improve and streamline the current manual quality assurance processes. We released three versions this fiscal year. The releases added new features to handle audiovisual (AV) files, serials, and the added logic to view PDF files instantly.

PHPartners & HSR Information Central Input System (HSRIC)

The Partners in Information Access for the Public Health Workforce Web site (PHPartners.org) was redesigned and implemented using responsive Web design methodology. This was a joint effort between two OCCS Applications teams and the LO Web and Information Management unit. Each PHPartners.org Topic page was enhanced with new updates and features.

We added new functionality to the PhPartners.org/HSR Information Central Input System by redesigning the menu tree page and enabling it to dynamically show multiple directories. The change will allow the directories to display more quickly so that users will find needed information much more efficiently.

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. During FY2014 over 1,900 new projects were added to HSRProj for searching. That increased the total number of projects available for searching to 28,450. We modified the data maintenance system login to use NIH credentials, increasing the security of the system and eliminating the need for users to have and maintain a separate login account. Also this year, Data Explorer (formerly Vivisimo) was implemented as the site's new search and retrieval system. It provides a better experience for users by clustering results and then allowing the search to be further refined by these clusters.

IT Infrastructure Services

NLM High Speed Communication Network Enhancements

In 2014, NLM continued to support and enhance its high speed communications network. The network is designed to be reliable and secure in order to support interconnection of NLM internal computing resources as well as high speed connections to external networks, namely, the NIH campus network, the NIH Consolidated Collocation Site (NCCS) in Sterling, VA, the public Internet, and the Internet2 national research network. These advanced network facilities enable access to, and delivery of NLM information services for staff, contractors, partners, and users. This state-of-the-art network enables NLM to fulfill its mission of collecting, organizing, and disseminating biomedical information throughout the world. In order to meet ever-increasing demands for new network services and higher bandwidth, OCCS continually enhances the network infrastructure and related services.

During the year, NLM continued a contract with Level3 in order to provide NLM's connection to the commodity Internet. The dedicated bandwidth of the NLM/Internet connection was three Gigabits per second (Gbps), with a capability to burst up to ten Gbps as needed. NLM's connection to the Internet2 network was upgraded from two 10Gbps links via the Mid Atlantic Exchange (MAX) at the University of Maryland, to a 100Gbps connection from Internet2 at Ashburn, VA, and a backup 10Gbps connection via the MAX. These increased network bandwidths enable timely exchange of large data sets between NLM and research partners around the world.

The network connection between NLM and the NIHNet was upgraded to a new 100Gbps link in coordination with the NIH Network Modernization Team. This will improve performance for NLM connections to NIHNet, the Trusted Internet Connection (TIC), and NLM/NCCS replication traffic. This connection is via the new 100Gbps CIT MPLS modernization network, although it is currently limited by a 10Gbps firewall at NIH.

Trusted Internet Connection (TIC)

The planned connections to the HHS Trusted Internet Connection (TIC) network were originally scheduled to take place in FY2013; however TIC network implementation was delayed until FY2014 by HHS. The NLM connection has been postponed until FY2015. The OMB has mandated that all federal agencies reduce the number of connections to the Internet and pass traffic through TIC approved security centers. HHS is providing its own TIC network and HHS security facilities. NLM obtained approval to pass "unrestricted" data traffic through its own established Internet and Internet2 connections, and only "restricted" traffic needs to pass through the HHS TIC network. This will allow NLM to continue to meet its mission in the most efficient and cost effective manner with respect to communications and network services.

Enabling IPv6

With the implementation of the latest Internet protocol, IPv6, NLM has the opportunity to communicate more effectively with the world wide health community, especially those who rely on IPv6 - as a result of the depleted address space of IPv4. OCCS continued the implementation of the Internet Protocol version 6 (IPv6) that runs as a dual stack (in parallel with the older IPv4 protocol).

In 2013, OCCS met the federal mandate to upgrade public/external facing servers and services (e.g. Web, email, DNS, ISP services, etc.) to use IPv6. In 2014, OCCS met the next IPv6 Federal mandate—to have internal client applications that communicate with public Internet servers and support enterprise networks to use

IPv6 by the end of FY2014. OCCS tested, then deployed IPv6 to all OCCS-supported network subnets.

This effort also included the selection and deployment of an IPAM (IP Address Management) solution for NLM. An IPAM system enables consistent management of both IPv4 and IPv6 IP addresses. It also provides for centralized DHCP and future management of DNS services within NLM. These systems make IPv4 and IPv6 network management tasks much easier and more streamlined.

Unified Communications Pilot

OCCS implemented a pilot project to evaluate the potential advantages to NLM that would be offered by features of Unified Communications (UC). 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) conferencing, and application and desktop sharing. VOIP implementation could reduce or eliminate antiquated and expensive POTS (Plain Old Telephone Service) networks, reducing costs. UC will also enable increased employee productivity. OCCS continues to work with NIH/CIT on plans for implementing UC at NLM.

NLM Data Center Enhancements—Uninterruptable Power Supply

All 480 batteries in the NLM Data center Uninterruptable Power Supply (UPS) system were replaced in order to insure ongoing reliability. The existing batteries had reached end-of-life, and during preventive maintenance we observed that multiple batteries were not meeting minimum standards. These failed batteries led to expensive incremental replacements of a handful of batteries every quarter, costing up to \$10 thousand per year. The costs would have only escalated as the original batteries aged further. The new batteries have a three year warranty, saving us on costly replacements in the near term, and overall, should last for five to seven years altogether with proper care and maintenance, allowing the UPS System to operate at an optimal level.

To better monitor the condition of our UPS system batteries, NLM acquired and commissioned a battery monitoring system to enable continuous real-time monitoring of the health of these batteries. The system monitors individual battery voltage, internal resistance, and temperature, the vital indicators of the health of a battery. These readings let us determine the health of batteries instantly, allowing the prompt replacement of defective batteries before they compromise system performance and reliability. Also, with this real-time monitoring in place, we have reduced routine preventive maintenance from quarterly to semi-annually, saving \$5,000 in recurring maintenance costs each year.

We also replaced a number of the capacitors in the same UPS system, as these devices had also reached

end-of-life. Filter capacitors are used to remove unwanted and undesired frequencies and electrical anomalies from the UPS system input and ensure a regulated output to the IT equipment. These capacitors have a life span of five to seven years and if not replaced, could cause the UPS system to malfunction and or fail catastrophically.

Redesign of Cabinet and Electrical Resources Return Savings

A comprehensive cost reduction analysis was performed on the facilities in use at the NIH Consolidated Collocation Site (NCCS) in Sterling, VA. As a result both cabinet space and electrical power services were realigned to minimize cabinet space and optimize electrical support. Cabinet space usage was reduced by 1,400 square feet, for a cost saving of 56 percent per month and electrical power circuits were reduced by 47 electrical circuits. As a result of these changes, NCCS monthly cost savings are 57 percent with no impact on services or cabinet power redundancy.

IT Security—Continuous Diagnostics and Mitigation (CDM) to confront emerging threats

Information security continues to evolve and confront emerging security threats. This year, the Heartbleed OpenSSL bug, and Internet Explorer ‘zero-day’ vulnerabilities surfaced suddenly and required the security team’s vigilant awareness and direction, and diligent follow-up actions from System Administrators to patch vulnerable systems ahead of any security intrusions. In addition, Web application attacks brought major data breach security incidents. In order to protect our digital assets against existing and new cyber-attacks that have the potential to impair our ability to provide critical information services, the NLM security team continued undertaking new security projects with preventive, protective, and responsive measures while focusing on application security. NLM has continuously strengthened its security posture and fully supports the Department of Homeland Security (DHS) Continuous Diagnostics and Mitigation (CDM) Program.

The CDM program is a holistic approach to fortifying the cybersecurity of computer systems and networks. It provides capabilities and tools that enable system and network administrators know the state of their respective systems and networks at any given time, and identify and rank problems for priority resolution. Continuous monitoring of networks for flaws and anomalies alerts administrators to attacks and intrusions, enabling faster responses to fix vulnerabilities.

In FY2014, CDM related activities include:

- Vulnerability management—the NLM conducts and contributes its credential vulnerability scan reports to NIH. The NIH Vulnerability Management Team then assembles all NIH Institute/Center (IC) credential scan results, and publishes monthly dashboards shared

among all ICs’ Chief Information Officer (CIO), and Information Systems Security Officers. We have continuously maintained an outstanding job in complying with these required credential scans on various operating systems. NLM attained the goal of the NIH 2014 approach—Tier 2 “Reduce targeted risk.” NLM has one of the highest success rates among all ICs currently being measured for successful credentialed scans, and reporting timely remediation on reported vulnerabilities.

- Patch Management and Security Configuration Management—NLM implemented the BigFix security and compliance suite provided by NIH, which will help discovery, patch management, security configuration management, and vulnerability management on client machines and servers. The tool also assists with software user analysis and inventory bundle.
- Consolidated Log Analysis—NLM has pulled more system logs into a centralizing log collection and management system, which helps to provide situational awareness through real-time event monitoring, alerting, and log analysis across platforms. The increased visibility to this vast amount of system and service log data empowers the security and system administrators to react to anomalies and malicious events and quickly analyze, and mitigate, any identify security and operational issues.
- Forensic and Incident Response—In collaboration with NIH, we deployed the Mandiant Intelligent Response (MIR) client throughout NLM on internal Windows workstations and public-facing Windows systems to improve incident response and forensics capabilities. This project helps us quickly identify any unknown compromised Windows desktops and servers. With NIH’s regular scanning, NLM has had no machines compromised since the commencement of the project.
- NLM works continuously with the HHS Computer Security Incident Response Center (CSIRC), which serves a primary role at the department level for all operational IT security. The FY2014 refresh of NLM-hosted hardware provides NLM and HHS CSIRC more processing power to perform network traffic analysis for IT operational support as well as network security investigation.

National Library of Medicine (NLM) Penetration Test/Network Assessment

The overall objective of a penetration test/network assessment is to conduct an internal system, network, and application vulnerability assessment to identify exploitable vulnerabilities on the internal network that exposes NLM to risks. This year the NLM security team again contracted for an independent, internal penetration test, assuring at least an annual independent assessment. While independent penetration tests have also been taken up as an

NIH-wide security program activity, they provide an independent test team and audits of each NIH Institute/Center only once every two years. These assessments help us advance our security posture and proactively identify any security weaknesses. Prompt remediation of identified vulnerabilities is further enforced with System Assessment and Authorization (SA&A) reviews for the system's Authority to Operate (ATO) required by Federal Information System Management Act (FISMA).

Application Security

Continuous Diagnostics and Mitigation targets also include addressing top application vulnerabilities on public facing Web sites. NLM must provide secure and trustworthy services for the general public by preventing application-level compromise from cyber-attacks that may damage our system and data integrity. NLM security administrators and software developers conduct both internal system scans and application scans to ensure there are no outstanding security issues for existing and new applications. The NLM ISSO also coordinated with the NIH security team to schedule external application scans of NLM's public-facing Web sites.

Websense Web Security with IPv6 Support

NLM implemented Websense Web Security that blocks Web threats to reduce malware infections, decrease security incidents, and meets HHS compliance requirements. In order to support the IPv6 mandate, NLM worked with the vendor to pilot, test, and implement the latest version of Websense appliance with IPv6 support in September 2014. As a result, NLM became an early adopter of the tool that can block malicious Web traffic from IPv6 Web Sites.

Desktop Security

For desktop computer security, NLM's automated patch management program applied over 130,000 patches on commodity desktops, fixing known vulnerabilities to software. More than 1,000 desktop computers were updated, with more than 24 million signatures; that is an 11 percent increase from FY2013 to demonstrate widespread viruses and continuous threats across cyberspace. The Desktop Security team continues to excel by ensuring that all OCCS-managed PCs are secure. The team was recognized once again by the NIH Penetration Test team for having zero vulnerabilities, now three years in a row.

Information Security Training

The Office of Personnel Management (OPM) requires annual security awareness for all staff that use IT resources. This training promotes awareness of the threats facing the NLM, and the best practices all staff can take to

counteract our security and privacy adversaries. NIH also merged the annual privacy awareness with security training this year. This requirement is a key activity reviewed by the HHS Inspector General (IG) during the annual FISMA audit. As in past years, NLM reached the targeted goal—to be the first IC to accomplish 100 percent completion of the mandate for both training courses this year. NIH also requires that all individuals identified as having "Significant IT Security Responsibilities" comply with the HHS requirement to successfully complete at least one relevant security training session every three years. This program targets executives, managers, and IT administrators for deeper insight into prevailing security threats and security practices. The NLM ISSO manages outreach and the validation of completion of this role-based training. This year, NLM included additional divisional directors and senior managers to take HHS Executive role-based training, so the total number of NLM staff to complete role-based training was doubled from that of last year.

Application Security is an essential part of the NLM Security program and it has a pivotal role in the success of NLM's many mission-critical applications. The NLM ISSO organizes and schedules two day, hands-on training sessions, and informs developers and managers of all divisions of upcoming sessions. These Web Application Security classes have received rave reviews from attendees.

NLM Disaster Recovery Test Schedules and Results

One of the requirements of the NIH Security Assessment and Authorization (SA&A) and Contingency Planning for continuity of NIH information technology (IT) services is for NLM to perform a "Disaster Recovery Test with Results" each year. The disaster recovery should occur either automatically or manually through the rerouting of service delivery over to the NIH Consolidated Collocation Site (NCCS) facility. The method used by each subsystem is based on business requirements as well as a cost/benefit analysis. The NLM ISSO worked with the point of contacts for NLM critical systems to follow systems designated disaster recovery (DR) methods and completed the test results. Test results were uploaded to the NIH Security and Accreditation Tool (NSAT).

Increase the Use of Virtualization and Green Computing Technology

OCCS is using virtualization to rapidly deploy information systems, boosting computing resource utilization, reducing Data Center footprint, and increasing the energy efficiency of NLM's mission critical applications. Virtualization is the foundation for our internal cloud strategy. During FY2014, OCCS increased the use of virtualization and green computing of server infrastructure by 12 percent. The number of virtualized servers remains about same as last year, but the number of physical servers has dropped

from 135 to 114. Eighty-two percent of all OCCS supported systems have been virtualized. This count is based on the number of virtual servers divided by the total

number of systems supported. The chart below shows the FY2014 Virtualization Progress through this fiscal year.

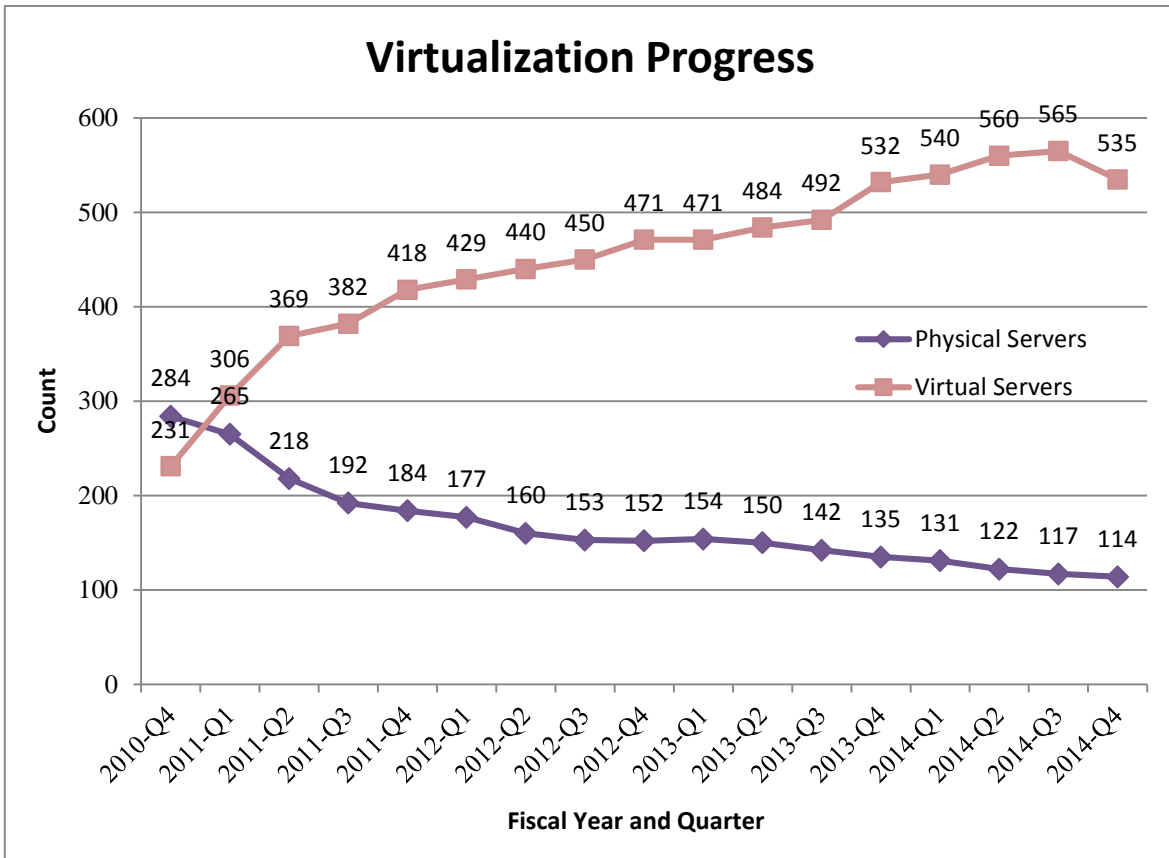


Figure 5. Virtualization Program Progress

We continue to move more standalone applications to virtual environments. As a result, computer rack usage dropped from 660 to 532, and power consumption dropped from 67 kW to 50 kW in the Data center during FY2014.

Improve The Efficiency and Effectiveness of Managed Resources by Reducing Duplicative Technologies or Systems

Where possible, OCCS is consolidating and eliminating duplicate technologies and systems. This reduces current maintenance and operational costs, but will also reduce the total cost of ownership of any new technology acquisition, including the resources required to manage the new technologies.

In FY2014, OCCS consolidated substantially within three different domains: older clustered enterprise storage systems, SecurID servers, and legacy hardware systems (Solaris). These consolidations resulted in

reduced operational and maintenance costs of \$50,000 per year, reducing the Data Center footprint, and decreased power and cooling load.

Enhance New Capabilities or Features:

OCCS is continually evaluating or implementing new capabilities or features for OCCS supported services to provide additional features, improve system performance, or enhance the end-user experience. In FY2014, we completed the following projects which provided new capabilities or features:

- Migrated the IHM Web site to newer hardware and OS which resulted in more rapid response times, from 200ms to 75ms.
- Upgraded the Citrix remote access service to a newer version of the software and hardware which provided additional features, added capacity, improved performance and enhanced end-user experience.

- Assisted with the implementation of the new CDE application technology at the disaster recovery site (NCCS) to provide fail-over capability in the event of disaster.
- Improved the availability and redundancy of RxImage API and SPLIMAGE Web sites by hosting in NLM Data Center. This permitted us to provide 24/7 support, reuse existing resources within the NLM Virtual Machine environment, and leverage other technologies/licenses.

Enterprise Computing and Storage Infrastructure Upgrade

We upgraded OCCS enterprise storage systems (NetApp) at the disaster recover site (NCCS) 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 added resiliency against peak demand. It also allows for future growth of data utilization. These storage systems currently manage 775TB raw storage. We will continue to evaluate the storage infrastructure for opportunities to consolidate and ways to increase efficiency.

Migration to the Windows 7 64-bit Operating System

Windows 7 performs more reliably and faster than the Windows XP. In FY2014, OCCS made unprecedented use of a method to build multiple PCs simultaneously in place. Migration of onsite and offsite commodity desktops was completed by the September 30, 2014 deadline, ahead of the planned start date to implement IPv6 on the desktop.

IT Web Platform Services

Web Analytics

NLM tracks 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.

We are shifting from Web log analysis to implementing a page-tagging methodology on 20 percent of our Web sites. Page tagging provides us a detailed view of how users navigate our Web sites and we can use this information to further improve our Web sites and increase customer satisfaction. Page tagging is more effective at incorporating the user's experience and perspective effectively, where Web log analysis is limited to actions seen at the Web server only. Additionally, page tagging reduces our licensing costs for Web traffic monitoring tools. We achieve these cost savings since with page tagging we no longer look at all traffic that hits our Web servers, particularly traffic from Web-crawling sites such Google, Bing, etc. Web monitoring tool license costs are proportionate to the size of Web logs, so smaller logs directly reduce license costs.

Database Platform Migration

OCCS strives to achieve increased database security, and to provide more efficient administrative processes. This year the new development/quality assurance (DEV/QA) Oracle RAC One Node Cluster was built on more powerful hardware, and databases were migrated to the new cluster. All OCCS managed databases have been upgraded. OCCS continued to proactively apply Oracle quarterly Critical Patch Updates (CPU) to the databases and support the systems group in applying OS security patches to the Oracle database servers. Oracle 12c database was installed for the developers to begin exploring new capabilities. A standby database was set up and tested for RxNorm at the NLM offsite disaster recovery center.

Medical Literature Support & Document Delivery Services

Data Creation and Maintenance System (DCMS)

OCCS achieved process improvements by updating the software used for maintaining MEDLINE. We automated structured abstract label discovery in NLM's Data Citation Maintenance System to discover structured abstract labels that otherwise would require substantial manual human effort.

The DCMS program automatically indexed 39,586 commentaries saving approximately \$396 thousand Indexing costs for FY2014.

DOCLINE®

We released version 5.0 of DOCLINE, NLM's information system that facilitates the routing of interlibrary loan requests. A key enhancement of this release is the enabling of interlibrary lending embargo periods within electronic journal serial holdings records. Libraries can now declare in advance whether titles are truly available for lending based on their licensing agreements with various publishers. The enhancement allows DOCLINE to bypass lenders who are unable to supply a requested article if the publication date falls within the publisher's embargo periods. This is expected to save on end-to-end processing time, and participating library staff time, by more accurately routing borrowing requests to libraries whose holdings are known to be immediately available for lending. At the end of the fiscal year, 86 libraries have set 4,475 embargoes for 2,687 electronic journals, and 563 Libraries have 185,882 linkout embargoes for 8,982 electronic journals.

Literature Selection Technical Review Committee (LSTRC)

The Literature Selection Technical Review Committee (LSTRC) is an advisory committee of external experts who recommend journal titles for indexing in MEDLINE®. The LSTRC administration system records current and

historical data concerning committee review of journal titles in the health and biomedical sciences.

Enhancements to the LSTRC administration system: improved publisher applications for inclusion in MEDLINE; streamlined selecting the review date for journals and assigning journals to the LSTRC committee; and automated the generation of the list of journals accepted for indexing. The combination of these improvements eliminate error-prone manual processes and save considerable Library Operations staff time. .

Search Support

In December 2013, the ENCompass software system was decommissioned, and NLM launched a new user interface using IBM InfoSphere Data Explorer (DE) to several key resources in the history of medicine: *Index-Catalogue, Library of the Surgeon-General's Office 1880-1961, A Catalogue of Incipits of Mediaeval Scientific Writings in Latin* by Thorndike and Kibre, and *A Catalogue of Incipits of Mediaeval Scientific Writings in Old and Middle English; an electronic resource* edited by Linda Voights and Patricia Kurtz. This implementation led to faster processing for internal staff; and for external users.

We released a new version of the Luna software, which supports access to Images from the History of Medicine (IHM). This improved the responsiveness of the Web site substantially.

We re-engineered the technology platform for the NLM Classification system, used by many libraries for arrangement of library materials in the field of medicine and related sciences. As a result of this initiative, online searches that used to take 20 seconds to complete are now completed in less than two seconds.

Outreach and Customer Services

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

Online Exhibitions

The Digital Gallery component of online exhibitions showcases curated collections of History of Medicine Division collections in a digitized format. Selected assets are drawn from the Digital Repository, the Images in the History of Medicine database, and newly digitized items

reflective of topics depicted in the online exhibitions and in support of education resources.

OCCS took the initiative to establish and implement a back-end database for the Digital Gallery component of online exhibitions showcases. Our long term goal for this project encompasses cost savings, creation of a database to support Web site content, and the ability to repurpose and reuse the Web site for multiple exhibitions. A database that can be repurposed for multiple exhibitions is an effective solution for today and an investment in our future. The first Digital Gallery was successfully released this year.

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

An outreach projects database is used internally to keep track of details about outreach activities, including who may receive direct benefits. We reengineered the Outreach Project Application (OPA) to improve performance. The new design produces search results in less than eight seconds. Users can now access both the OPA and the Web exhibits application in a single login. We also improved the reporting functionality, which can now produce reports in HTML, Excel, and PDF formats.

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 data on the use of NLM’s Internet1 and Internet2 data communications links are also displayed.

OCCS implemented new features and upgrades to the NPDS. We replaced one map with three interactive temporal maps, updates to maps, addition of a word cloud, and implemented 16 page view maps. Interactive maps allow customers and library patrons to visualize and analyze Web site traffic-places and dates of visits and searches by time.

ADMINISTRATION

Todd D. Danielson
Associate Director for Administrative Management

Table 8. Financial Resources and Allocations, FY2014
(Dollars in Thousands)

Budget Allocation	Amount
Extramural Programs	\$60,413
Intramural Programs	<u>\$254,018</u>
Library Operations	\$(63,137)
Computer and Communications Systems	\$(31,386)
Lister Hill National Center for Biomedical Communications	\$(44,220)
National Center for Biotechnology Information	\$(95,838)
Specialized Information Services	\$(19,437)
Research Management and Support	<u>\$13,982</u>
Total Appropriation	\$328,413
Plus: Reimbursements	<u>\$58,721</u>
Total Resources	\$387,134

Personnel

Table 9. FY2014 Full-Time Equivalent (Actual)

Program Area	No.
Office of the Director	14
Office of Health Information Programs Development	5
Office of Communication and Public Liaison	10
Office of Administration	52
Office of Computer and Communications Systems	46
Extramural Programs	22
Lister Hill National Center for Biomedical Communications	69
National Center for Biotechnology Information	283
Specialized Information Services	44
Library Operations	<u>254</u>
TOTAL FTEs	799

New Appointments

In October 2013, **Sara Tybaert** was appointed as the Head of the MEDLARS Management Section (MMS) in Library Operations. Ms. Tybaert has been with NLM since 1990 filling many roles during that time. Since 2011, she has been the Head of the Bibliographic Data Management Unit of MMS. Ms. Tybaert received an MLS from Florida State University and a BA from the University of Maryland.

In January 2014, **Rebecca Warlow** was appointed as Head of the Images & Archives Section in the NLM's History of

Medicine Division. She came to this position with over 15 years of experience at the National Archives and Records Administration (NARA), where most recently she was the Digitization Partnerships Coordinator in NARA's Office of Innovation. Ms. Warlow has a master's degree in history from Northeastern University and a bachelor's in history from Dickinson College.

In February 2014, **Mark E. de Jong** was appointed as Head, Collection Access Section, Public Services Division of Library Operations. Mr. de Jong came to NLM from the University of Maryland University College (UMUC) where he was the Access Services Manager and directed document delivery and interlibrary loan for UMUC's students and faculty in the US, Europe, and Asia. He was an adjunct associate professor at UMUC where he taught both undergraduate and graduate courses in research and information literacy. Shortly after arriving at NLM, he received a doctorate in management from UMUC

In April 2014, **Benjamin E. Petersen** was appointed as the Head, Preservation and Collection Management Section, in the Public Services Division of Library Operations. Mr. Petersen comes to NLM from the Southern History Department at the Birmingham Public Library, where he had been Head since 2008. At Birmingham Public, he managed operations and programming for a well-known collection of southern history and genealogy. He is an expert in environmental monitoring systems, and has advised other libraries in that area. Mr. Petersen holds a joint master's degree in public history and library and information science from the University of South Carolina in Columbia.

In June 2014, **Dar-Ning Kung, PhD**, was appointed to the position of the Associate Director for Information Assurance, Office of Computer and Communications Systems. Dr. Kung joined NLM in October 2001, and has been serving as the Information Systems Security Officer (ISSO) for NLM since 2002. In the role of the NLM ISSO, he has been instrumental in advancing NLM's IT security posture through the implementation of many key security initiatives. In his new role, Dr. Kung will be responsible for developing strategies for protecting the confidentiality, integrity and availability of NLM's digital resources. He will continue to serve as NLM's ISSO and represent NLM's security interests within NIH and HHS. He has served as President, as well as other official positions, of the Asian Pacific Islander Employment Committee (APIEC) and the NIH Asian Pacific American Organization (APAO), for several years. He received his PhD in Mechanical Engineering from the University of Maryland, College Park; his MS in computer science from Johns Hopkins University in 1994; his MS in mechanical engineering from University of Maryland, College Park; and his BS in mechanical engineering from National Cheng Kung University, Taiwan, ROC.

NLM Associate Fellows Program for 2014–2015

Ariel Deardorff received her MLIS in 2014 from the University of British Columbia in Vancouver, Canada. As an MLIS student, Ms. Deardorff worked at the University of British Columbia's Education Library, Office of Library Assessment, and Faculty of Medicine, and volunteered at the British Columbia Children's Hospital Library. During her MLIS program, Ms. Deardorff was active in her school's chapter of the British Columbia Library Association, and the Special Library Association, and served as a student representative to the group Academic Librarians in Public Service. In 2013 Ms. Deardorff was chosen to represent her school as part of the American Library Association's Student to Staff program, through which she was able to volunteer at the 2013 ALA annual conference. Ms. Deardorff holds a bachelor's degree in international studies from the University of Washington.

Kristina Elliott received her MLS from the University of Maryland in 2014. While completing her degree, Ms. Elliott worked as a graduate assistant in the College of Information Studies, supporting communications, marketing, and alumni outreach efforts. As a master's student, she interned with the National Library of Medicine, developing a social media strategy and guidelines for the Library's Twitter and Facebook accounts. She was also a metadata and social media intern with NOAA Central Library in Silver Spring, MD. Before beginning her master's program, Ms. Elliott worked for a cloud-based marketing software company as an editor, writer, and researcher. She earned her BA in English from the University of Maryland, College Park.

Erin Foster received her MSLS degree in 2014 from the University of North Carolina, Chapel Hill. While completing her degree, Ms. Foster worked as a graduate research assistant at UNC Chapel Hill's Health Sciences Library, providing reference and instruction services, as well as faculty and student research support. Ms. Foster also worked on several projects with UNC Chapel Hill's digital repository and completed a field experience at Duke University in Rubenstein Library's Technical Services Department. Prior to graduate studies, Ms. Foster worked as a library assistant and data manager at an aerospace and defense company. Ms. Foster completed her undergraduate degree at the University of California, Santa Cruz, where she double majored in literature and feminist studies. As an undergraduate, Ms. Foster held student assistant positions in the Reference and Media Center areas of UC Santa Cruz's McHenry Library.

Lori E. Harris received her MSLS from the University of North Carolina at Chapel Hill. Ms. Harris was a recipient of the American Library Association Spectrum Scholarship in 2012-2013 and the Society of American Archivists' Harold T. Pinkett Minority Student Award in 2013-2014. As a graduate student, Ms. Harris worked in

North Carolina State University's Special Collections Research Center as both a Library Associate and Project Archivist. Ms. Harris' interest in medical librarianship was sparked when she worked as an archivist intern at the Black Women's Health Imperative in Washington, DC. Ms. Harris holds a BA in American Studies and two certificates in Archival Studies and Museum Studies, from Smith College.

Departures

In April 2014, **Mehryar Ebrahimi** left NLM to assume the position as Head the FDA's Design and Engineering Branch. Mr. Ebrahimi joined the NLM in November 2004, as the Director of OAMAS. In this capacity, he was the caretaker of the NLM infrastructure. Over the years, he has spearheaded a number of initiatives that brought the NLM facilities into compliance with modern engineering standards and vastly improved space utilization.

Awards

The 2014 Paul Evan Peters Award was presented to **Dr. Donald A.B. Lindberg** by the Coalition for Networked Information (CNI), the Association of Research Libraries (ARL), and EDUCAUSE, in recognition of his direction, sustained visionary and high-impact leadership in using networked information to transform everything from consumer health care to fundamental research in molecular biology and related disciplines.

The 2013 CENDI Meritorious Service Award was presented to **Jerry Sheehan**, Assistant Director for Policy Development, National Library of Medicine, by CENDI, the federal Scientific and Technical Information Managers Group for his significant contributions to CENDI's information policy programs and discussions over several years.

The Thomson Reuters/Frank Bradway Rogers Information Advancement Award was presented to the MedlinePlus Connect Group, including **Joyce E.B. Backus, Stephanie N. Dennis, Jennifer M. Jentsch, Naomi Miller, Donald J. Potvin, Jr., Dr. Rex R. Robison, Somanath V. Lanka** (contractor), **Dr. Maxine L. Rockoff** (Columbia University), **Dr. Neil Calman** (Institute for Family Health NYC), and **Dr. Stephen H. Bryant** (NYU Langone Medical Center) by the Medical Library Association in recognition of outstanding contributions for the application of technology to the delivery of health science information, to the science of information, or to the facilitation of the delivery of health science information.

The HHS recipient of The Federal Engineer of the Year Award for 2014 was **Mehryar Ebrahimi**, Director, Office of Administrative and Analysis Management Services (OAMAS), in recognition of his use of new IT technologies coupled with innovative application of

mechanical and electrical systems resulting in a data center that meets the growth requirements as well as reduction in energy consumption.

The Hubert H. Humphrey Award For Service To America was presented to **Dr. Terry S. Yoo** at the 2013 HHS Departmental Awards for leadership in open science through Insight Toolkit (ITK) as the public software standard for three-dimensional biomedical image analysis and for all the patients who have benefited.

NCBI Director **David Lipman**, MD, received the Jim Gray eScience Award from Microsoft Research. The award recognizes researchers who have made outstanding contributions to the field of data-intensive computing in the pursuit of open, supportive, and collaborative research models.

The HHS*ignite* Pathway Award was presented to an NLM team including **Dr. Sameer K. Antani**, **Dr. Stefan R. Jaeger**, **Dr. Alexandros Karargyris**, **Dr. George R. Thoma**, **Dr. Sema Candemir** (contractor), and **Dr. Zhiyun Xue** (contractor) for their project, Automatic X-Ray Screening for Rural Areas. Imaging scientists have developed an algorithm to auto-detect TB in chest x-rays in order to meet radiologist shortage in rural areas. The algorithms have been validated, allowing radiologists to view and confirm only the x-rays that are identified (by the software) as exhibiting the disease. This team proposed to develop a user-friendly interface for their application so that the technique can be tested by and optimized for providers at the point of care.

The NLM Board of Regents Award for Scholarship or Technical Achievement is awarded to recognize and stimulate independent creativity leading to scholarly and/or technical achievements that enrich biomedicine. The recipient of the 2014 award was **Dr. Olivier Bodenreider**, in recognition of his exceptional leadership in the dissemination of NLM drug information sources through graphical and programming interfaces.

The Frank B. Rogers Award recognizes employees who have made significant contributions to the Library's fundamental operation programs and services. The recipient of the 2014 award was **Janet R. Zipser** for outstanding leadership in reinventing the NLM training program, incorporating new technologies, adult education practices, and user feedback.

The NLM Director's Honor Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to **Dr. Barbara A. Rapp**, for exceptional contributions as NLM's Planning and Evaluation Officer, synthesizing information and formally reporting progress of NLM toward meeting key goals of its mission.

The EEO Special Achievement Awards recognize employees for their efforts to advance equal employment opportunities at NIH and their ability to motivate others. The Phillip C. Coleman Award was presented to **Dr. Kathel Dunn** and **Lou Wave S. Knecht** for leadership and contributions in the development and management of the Library Operations (LO) Career Enrichment Program. The EEO Special Achievement Award was presented to **Sally S. Boakye** for outstanding leadership, management, and mentoring skill in assisting employees to achieve their highest career potential.

The NIH Merit Award recognizes individuals or groups whose superior service and achievement warrant special appreciation. The NIH Merit Award was presented to seven individuals including: **Deena S. Acton** for exceptional contributions in managing Voyager reporting functions for Library Operations staff to support budget management, workflow analysis, monitoring of production goals, and enhanced data quality for NLM users; **Roger W. Brtva** for creativity in improving NLM's systems and supporting customer service, resulting in increased efficiency of information delivery to staff and the public; **Dora A. Deegbe** for unparalleled contributions to Library Operations and NLM administration in managing human resources, recruitments, training, and diversity initiatives; **Michael J. Gill** for leading the development and launch of NLM's Family Reunification systems, both for community-wide as well as for hospital-centered disaster events; **Shannon M. Jordan** for leadership and direction in coordinating NLM's complex Hazardous Substances Database and administering a reinvention process to make the creation and use of this valuable resource more effective and efficient; **James B. Labosier** for outstanding service to the NLM, its patrons, and the archival community through his commitment to excellence in archival processing, access, and care of collections; and **Edwin P. Sequeira** for his creativity, expertise, and exceptional efforts in fostering the development of NIH PubMed Central.

The NIH Director's Award was presented to two NLM staff member as individual awards and four NLM staff as part of NIH group awards. Individual awards were received by **Dr. Dar-Ning Kung** in recognition of contributions to advancing information technology and leadership in establishing the OMB-mandated Trusted Internet Connection (TIC) at the NLM, and **Jerry R. Sheehan** for outstanding achievement as principal drafter and point person for the development of the Notice of Proposed Rulemaking for Clinical Trials Registration and Results Submission. The Office of the Director, NIH, HeLa Whole Genome Sequence Data Group Award included **Michael L. Feolo** and **Dr. James M. Ostell** for exceptional work on the historic agreement with the Henrietta Lacks family that allows biomedical researchers to access HeLa whole genome sequence data. The National Institute of Arthritis and Musculoskeletal and Skin

Disease, Trans-NIH American Indian/Alaska Native Team Group Award included **Laura L. Bartlett** and **Gale A. Dutcher** in recognition of outstanding dedication and sustained commitment to improving NIH's ability to effectively reach American Indian and Alaska Native communities. The Common Fund Leadership Award was presented to **Dr. Stephen H. Bryant** as part of Office of the Director, NIH, NIH Molecular Libraries Program Group Award for extraordinary scientific leadership and management of the NIH Molecular Libraries Program as a resource to enable discovery of small molecule probes.

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 FY2014, the fund approved 28 staff to take 37 courses, the majority of which were undergraduate. The school with the largest number of NLM enrollees was the University of Maryland (20). Other institutions included Strayer, American University, Catholic University, Kent State, Mount St. Mary's, USDA Graduate School, and Syracuse University. Course disciplines included: Computer Science, Management, Law, Public Health, Forensics, Economics, English, Library Science, Physics and Language among others. Less than 10 percent were online courses. The Diversity Council continues its efforts to publicize the availability of the fund.

NLM 2014 Health Exposition

The NLM Diversity Council sponsored the seventh annual Healthy Lifestyles for You and Your Family and Scientific Expo on April 24, 2014. NLM's educational family-oriented health expo focusing on healthy lifestyles for both adults and children and scientific awareness was attended by approximately 1,000 children and their parents. Activities included Mad Science, innovative, hands-on Harnessing Heat and Watt's Up workshops for children to encourage scientific literacy in children and Operation Lunch Line 3D, presented by Kid Power's Programs. Harnessing Heat was a scientific experiment in which the children built a thermometer; watched an underwater volcano erupt and changed the states of matter. Watt's Up taught the children about electricity and its properties. The children used experiments to make indoor thunder and lightning. Operation Lunch Line 3D was a highly interactive, musical show, designed to help children learn the value of good nutrition and exercise. Using spectacular visual effects in 3D, the entire audience miniaturizes, joining Kid Power on an amazing journey inside the human body of a boy named Max who feels lousy because he doesn't eat or move properly. Through audience participation the kids not only educate and motivate Max, but in doing so, learn they too are special, filled with all the "kid power" needed to develop the knowledge, skills, attitudes and behaviors necessary to make health-enhancing choices and feel great. The NLM Healthy Lifestyles and Scientific Expo has received positive recognition from parents, other institutes and the public in promoting healthy lifestyles and scientific education to children in their formative years.

Appendix 6: 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 John Freeman Blvd.
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
12-077 Center for the Health Sciences
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
55 Lake Avenue North Rm S4-241
Worcester, MA 01655-0002
Phone: (508) 856-5979 Fax: (508) 856-5977
States served: CT, MA, ME, NH, RI, VT
URL: <http://nnlm.gov/ner>

Appendix 7: 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.

Asterisk () denotes members of the BOR EP Subcommittee*

Chairperson

MacKay, Trudy, PhD* (Subcommittee Chair)
Distinguished University Professor of Genetics
Department of Genetics
North Carolina State University
Raleigh, NC 27695-7614

Appointed Members

Dishman, Eric, MS
Intel Fellow and General Manager
Intel Corporation
Hillsboro, OR 97124

Fleming, David A, MD
Professor and Chairman
Department of Internal Medicine
Director, MU Center for Health Ethics
University of Missouri School of Medicine
Columbus, MO 65212

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
Miramar, FL 33029

Martin, Sandra, MSLS
Director
Shiffman Medical Library
Wayne State University
Detroit, MI 48201

Roskies, Ralph Z, PhD*
Professor of Physics, University of Pittsburgh
Scientific Director, Pittsburgh Supercomputing Center
Pittsburgh, PA 15213

Sternberg, Esther M, MD
Professor of Medicine
University of Arizona
College of Medicine
Tucson, AZ 85724

Yokote, Gail A, MS*
Associate University Librarian for Sciences and Technical Services
Peter J. Shields Library
University of California, Davis
Davis, CA 95616

Ex Officio Members

Billington, James H, DPhil
Librarian of Congress
Library of Congress
Washington, DC 20540

Collins, James, PhD
Assistant Director, Biological Sciences
National Science Foundation
Arlington, VA 22230

Horoho, Patricia D, BSC, MSN
The Surgeon General/Commander
US Army Medical Command
Falls Church, VA 22041-3258

Liu, Simon Y, PhD
Director
National Agriculture Library
U.S. Department of Agriculture
Beltsville, MD 20705-2351

Lushniak, Boris D, RADM
Acting Surgeon General
Office of the Surgeon General
Office of the Assistant Secretary for Health
Washington, DC 20201

Nathan, Matthew L, BS
Surgeon General of the Navy
Chief, Bureau of Medicine and Surgery
Department of the Navy
Washington, DC 20372-5300

Clancy, Carolyn M, MD
Acting Under Secretary for Health
Veterans Health Administration
Washington, DC 20420

Rice, Charles L, MD
President
Uniformed Services University of the Health Sciences

Bethesda, MD 20814-4799

Travis, Tom, Major General, USAF, MC
Surgeon General
United States Air Force
Washington, DC 22209

Appendix 8: 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

Member

Berner Weiss, Eta S., EdD
Professor, Health Informatics
Department of Health Services Administration
School of Health Professions
School of Medicine
University of Alabama at Birmingham
Birmingham, AL 35294-3361

Cummins, Mollie R., PhD
Associate Professor
School of Nursing and Medicine
University of Utah
Salt Lake City, UT 84112-5880

Hammond, William E., PhD
Director

Duke Center for Health Informatics
Duke University
Durham, NC 27705

Hicks, LeRoi S., MD
Vice-President of Medicine
Department of Medicine
Christiana Care Health System
Newark, DE 19718

Rucker, Donald W., MD
Chief Operating Officer, IDEA Studio
Ohio State University Medical Center
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

Xu, Hua, PhD
Associate Professor
School of Biomedical Informatics
The University of Texas Health Science Center
Houston, TX 77030

Appendix 9: 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

Boehnke, Michael L., PhD
Professor
Department of Biostatistics
School of Public Health
University of Michigan
Ann Arbor, MI 48109

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

Green, Rachel, PhD
Department of Molecular Biology and Genetics
School of Medicine
John Hopkins University
Baltimore, MD 21205

Guy, R. Kiplin, PhD
Professor
Department of Chemical Biology and
Therapeutics
St. Jude Children's Research Hospital
Memphis, TN 38105

Wu, Chung-I, PhD
Professor
Department of Ecology and Evolution
University of Chicago
Chicago, IL 60637

Zhang, Jianzhi, PhD
Marshall W. Nirenberg Collegiate Professor
Department of Ecology and Evolutionary Biology
University of Michigan
Ann Arbor, MI 48109

Appendix 10: 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

Lussier, Yves A, MD
Professor of Medicine and Engineering
Clinical Research Information Officer
Assistant Vice President of Health Affairs
The University of Illinois in Chicago
Chicago, IL 60612

Members

Alpi, Kristine M, MLS, MPH
Director and Adjunct Assistant Professor
Department of Population Health & Pathobiology
Veterinary Medicine Library
North Carolina State University
Raleigh, NC 27607

Archer, Kellie J, PhD
Associate Professor
Department of Biostatistics
Virginia Commonwealth University
Richmond, VA 23298-0032

Bennett, Kristin P, PhD
Professor
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY 12180

Butte, Atul J, MD, PhD
Associate Professor
Systems Medicine/Pediatrics
Stanford University
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 10014

Craven, Mark W, PhD
Professor
Department of Biostatistics and Medical Informatics
University of Wisconsin-Madison
Madison, WI 53706

Denny, Joshua C, MD
Associate Professor
Departments of Medicine and Biomedical Information
Vanderbilt University
Nashville, TN 37232

Holmes, Kristi, PhD
Director
Galter Health Sciences Library
Feinberg School of Medicine
Northwestern University
Chicago, IL 60611

Kho, Abel N, MD
Assistant Professor
Division of General Internal Medicine & Biomedical Informatics
Northwestern University
Chicago, IL 60611

Koonce, Taneya Y, MPH, MSLS
Deputy Director
Knowledge Management/Eskind Biomedical Library
Vanderbilt University Medical Center
Nashville, TN 37232-8340

Larson, Ray R, PhD
Professor
School of Information
University of California, Berkeley
Berkeley, CA 94720

Mooney, Sean D, PhD
Associate Professor
Department of Biomedical Informatics
Buck Institute for Research on Aging
Novato, CA 94945

Payne, Philip, PhD
Director, OSU Data Analytics Collaborative
Chair, Department of Biomedical Informatics
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
Director
Ruth Lilly Medical Library
Indiana University School of Medicine
Indianapolis, IN 46202-5121

Savova, Guergana K, PhD
Associate 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

Strayhorn, Gregory, MD, PhD
Professor
Department of Family Medicine
Morehouse School of Medicine
East Point, GA 30344

Szolovits, Peter, PhD
Professor
Computer Science and Artificial Intelligence Lab
Massachusetts Institute of Technology
Cambridge, MA 02139

Weng, Chunhua, PhD
Associate Professor
Department of Biomedical Informatics
Columbia University
New York, NY 10032

Appendix 11: 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

Pascoe, John M., PhD
Professor
Department of Pediatrics
Wright State University
Dayton, OH 45404

Members

Cabello, Felipe C., MD
Professor
Department of Microbiology & Immunology
New York Medical College
Valhalla, NY 10595

Cheung, Dorothy S., MD
Assistant Medical Director
Immunology, Tissue Growth and Repair
Early Clinical Development
Genentech, Inc.
South San Francisco, CA 94080-4990

Corsi, Karen F., ScD
Associate Professor
Department of Psychiatry
University of Colorado, Denver
Denver, CO 80206

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

Nwomeh, Benedict C., MD
Attending Surgeon, Professor
Pediatric Surgery
Nationwide Children's Hospital
Ohio State University
Columbus, OH 43205

Pascoe, John R., BVSc, PhD
Professor and Executive Associate Dean
Dean's Office
School of Veterinary Medicine
University of California, Davis
Davis, CA 95616

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

Yoshimura, Masami, DSc
Associate Professor
Department of Comparative Biomedical
Sciences
School of Veterinary Medicine
Louisiana State University
Barton Rouge, LA 70803

Appendix 12: 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., M.L.S., M.B.A.
Associate Dean for Library Services & Archives
Medical Center Library
Duke University
Durham, NC 27710

Members

Bedard, Martha A., MSLS, MA
Vice Provost for University Libraries
University of Connecticut
Storrs, CT 06269

Cantu, Adelita G., PhD
Assistant Professor
Family & Community Health Systems
The University of Texas Health Science Center
San Antonio, TX 78229

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

Dewey, Barbara, MD
Dean
University Libraries and Scholarly Communications
Pennsylvania State University
University Park, PA 16802-1812

Eisen, Jonathan A., PhD
Professor
Departments of Medical Microbiology and Immunology
UC Davis Genome Center
University of California, Davis

Davis, CA 95616

Engelward, Bevin P., PhD, ScD
Associate Professor
Department of Biological Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Haricombe, Lorraine J., PhD
Dean of Libraries
Watson Library
University of Kansas
Lawrence, KS 66045

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

McCrary, Jr., Victor
Vice President for Research and Economic Development
Morgan State University
Baltimore, MD 21251

Morse, Randall H., PhD
Research Scientist
Molecular Genetics Program
Wadsworth Center
Albany, NY 12201

Stodden, Victoria, PhD
Assistant Professor of Statistics
Department of Statistics
Columbia University
New York, NY 10027

Terry, Sharon F., MA
President and CEO
Genetic Alliance
Washington, DC 20008-2304

Appendix 13: Organizational Acronyms and Initialisms Used in this Report

<u>Acronym</u>	<u>Meaning of Acronym</u>	<u>Acronym</u>	<u>Meaning of Acronym</u>
AAHSL	Association of Academic Health Sciences Libraries	BCALA	Black Caucus of the American Library Association
AABB	Non-profit association formerly known as American Association of Blood Banks	BD2K	Big Data to Knowledge initiative (NIH)
AAPA	American Academy of Physicians Assistants	BGMUT	Blood Group Antigen Gene Mutation Database
ABC	Advanced Biomedical Tele-Collaboration Test Bed)	BHEPP	Bethesda Hospitals' Emergency Preparedness Partnership
ACIOP	AIDS Community Information Outreach Program	BISTI	Biomedical Information Science and Technology Initiative
ACLA	American Clinical Laboratory Association	BITA	Biomedical Image Transmission via Advanced Networks
ACORN	Automatically Creating OLDMEDLINE Records for NLM	BLAST	Basic Local Alignment Search Tool
ACP	American College of Physicians	BLIRC	Biomedical Library and Informatics Review Committee
ACSI	American Customer Satisfaction Index	BMT	Boundary Marking Tool
AFIP	Armed Forces Institute of Pathology	BN	Brand Name
AG	Access Grid	BOR	Board of Regents
AHCJ	Association of Health Care Journalists	BSAT	BMT Study Administration Tool
AHIC	American Health Information Community	BoSC	Board of Scientific Counselors
AHILA	Association for Health Information and Libraries in Africa	BSD	Bibliographic Services Division
AHRQ	Agency for Healthcare Research and Quality	BSE	Book Scan Editor
AIDS <i>info</i>	Acquired Immune Deficiency Syndrome <i>info</i> (database)	BSN	Bioinformatics Support Network
AJPP	African Medical Journal Editors Partnership Program	C3PI	Computational Photography Project for Pill Identification
ALTBIB	Alternatives to Animal Testing	CADTH	Canadian Agency for Drugs and Technologies in Health
ALTS	ASCUS-LSIL Triage Study	CAM	Complementary and Alternative Medicine
AIH	American Indian Health (Web portal)	C&A	Certification & Accreditation (audit)
AME	Automated Metadata Extraction	CARE	Candidate Gene Association Resource project
AMIA	American Medical Informatics Association	CAS	Collection Access Section
AMPA	American Medical Publishers Association	CBB	Computational Biology Branch
AMPATH	Academic Model Providing Access to Healthcare	CBIR	Content-Based Image Retrieval
AMWA	American Medical Women's Association	CBRNE	chemical, biological, radiological and nuclear (incidents)
APDB	Audiovisual Program Development Branch	CCB	Configuration Control Board
API	Applied Programming Interface	CCDS	Consensus CoDing Sequence
APIRE	American Psychiatric Institute for Research and Education	CCHD	Critical Congenital Heart Disease
AREA	Academic Research Enhancement Award	CCHIT	Commission for Healthcare Information Technology
ARL	Association of Research Libraries	CCR	Central Contractor Registration
ARRA	American Recovery and Reinvestment Act	CCRIS	Chemical Carcinogenesis Research Information System
ASCCP	American Society for Cervical Pathology and Colposcopy	CCW	CMS Chronic Conditions Data Warehouse
ASHG	American Society of Human Genetics	CDART	Conserved Domain Architecture Retrieval Tool
ASPE	Assistant Secretary for Planning and Evaluation, HHS Office of the	CDD	Conserved Domain Database
ASPR	Assistant Secretary for Preparedness and Response, HHS Office of the	CDE	Common Data Element Resource Portal (NIH)
ATC	Anatomic Therapeutic Class		
ATO	Authority to Operate		
BAC	Bacterial Artificial Chromosome		
BarSTool	Barcode Submission Tool		

<u>Acronym</u>	<u>Meaning of Acronym</u>
CDM	Continuous Diagnostics and Mitigation Program
cDNA	Complementary DNA
CEB	Communications Engineering Branch
CEL	Affymetrix Cell intensity (file)
CEP	Career Enhancement Program (ARL)
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
CHIQA	Consumer Health Information and Question Answering (system)
CHRIS	Consumer Health Resource Information Service
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
CPU	Critical Patch Updates
CRAC	Computer Room Air Conditioner
CRAH	Computer Room Air Handler
CRD	Centre for Reviews and Dissemination (England)
CRI	Clinical Research Informatics
CRIS	Clinical Center EMR System (NIH)
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

<u>Acronym</u>	<u>Meaning of Acronym</u>
CTD	Comparative Toxicogenomics Database
CTS	Communications Technology Satellite
CTS2	Common Terminology Services 2
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
DDICC	Data Discovery Index Coordination Consortium
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
DIC	Disseminated Intravascular Coagulation
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
DR	Disaster recovery
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

<u>Acronym</u>	<u>Meaning of Acronym</u>
dW	docWorks (software)
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
EMSC	Emergency Medical Services for Children
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
EUREKA	Exceptional, Unconventional Research Enabling Knowledge Acceleration
E-Utilities	Entrepreneurial 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
FOA	Funding opportunity announcements (NIH)
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

<u>Acronym</u>	<u>Meaning of Acronym</u>
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
HITECH	Health Insurance Technology for Economic and Clinical Health
HITSP	Healthcare Information Technology Standards Panel
HLA	Human Leukocyte Antigen
HL7	Health Leven Seven, Inc.
HMD	History of Medicine Division
HSDB	Hazardous Substances Data Bank
HPCC	High Performance Computing and Communications
HPV	Human Papillomavirus
HRSA	Health Resources and Services Administration
HSRIC	HRS (Health Services Research) Information Central
HRSInfo	Health Services Research Information
HSRProj	Health Services Research Projects
HSRR	Health Services and Sciences Research Resources
HSTAT	Health Services and Technology Assessment Text
HTTP	Hypertext Transfer Protocol
HuGENet	Human Genome Epidemiology Network
HUMLO	Hemoglobinopathies Uniform Medical Language Ontology Project (NHLBI)
i2b2	Informatics Integrating the Bench and Bedside
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

<u>Acronym</u>	<u>Meaning of Acronym</u>	<u>Acronym</u>	<u>Meaning of Acronym</u>
ICD	International Classification of Diseases	JDI	Journal Descriptor Indexing
ICMJE	International Committee of Medical Journal Editors	JDMS	Journal Descriptor Maintenance System
ICs	Institutes and Centers (of NIH)	JRE	Java Runtime Environment
ICT	Information and Communication Technologies	KEGG	Kyoto Encyclopedia of Genes and Genomes
ICU	Intensive Care Unit	KSS	Knowledge Source Server (data)
IDE	Integrated Development Environment	LactMed	Drugs and Lactation (database)
IDS	Intrusion Detection System	LAN	Local Area Network
IE8	Internet Explorer 8	LANDHI	Latin American Disaster and Health Information Network
IEB	Information Engineering Branch	LC	Library of Congress
IGS	Intergenic Spacer	LHI	Leading Health Indicators (HHS)
IGSTK	Image Guided Surgery Toolkit	LHNCBC	Lister Hill National Center for Biomedical Communications
IHTSDO	International Health Terminology Standards Development Organization	LID	Laboratory for Informatics Development
IHM	Images from the History of Medicine	LIPID MAPS	Lipid Metabolites and Pathways Strategy
II	Indexing Initiative	LITE	Librarian Infobutton Tailoring Environment
ILL	Interlibrary Loan	LJI	List of Journals Indexed
ILS	Integrated Library System	LO	Library Operations
ImageCLEF	Image Cross Language Evaluation Forum	LOINC	Logical Observations Identifiers, Names, Codes
I-MAGIC	Interactive Map-Assisted Generation of ICD Codes	LPF	Lost Person Finder
IMPAC	Information Management Planning Analysis And Coordination	LRG	Locus Reference Genomic
IMPPOA	In-Memory Processing for Publisher Online Articles	LRP	Long Range Plan (NLM)
InCHIs	IUPAC International Identifiers	LSD	Lysosomal Storage Disorders
INDSC	International Nucleotide Sequence Database Collaboration (formerly DDBJ/EMBL/GenBank)	LSI	List of Serials Indexed
<i>infoSIDA</i>	<i>info</i> Síndrome de Inmunodeficiencia Adquirida (database)	LSTRC	Literature Selection Technical Review Committee
INN	International Nonproprietary Names	LVG	Lexical Variant Generator
IOC	Indicator of Compromise	LWS	Lifecycle Work Station
IP	Interactive Publications	MARC	Machine-Readable Cataloging
IPAM	IP Address Management	MARG	Medical Article Records Groundtruth
IPv6	Next Generation Internet, Version 6	MARS	Medical Article Records System
IRB	Institutional Review Board	MAX	Mid Atlantic Exchange, U. of Maryland
IRC	In-Row Coolers	MCI	Mass Casualty Incident
IRIS	Integrated Risk Information System	MDoT	MEDLINE Database on Tap
IRMA	Image Retrieval for Medical Applications	MDT	Multimedia Database Tool
ISO	International Organization for Standardization	MEDLARS	Medical Literature Analysis and Retrieval System
ISTO	Image Storage and Transmission Optimization	MEDLINE	MEDLARS Online
IT	Information Technology	MegaBLAST	Basic Local Alignment Search Tool
ITP	Informatics Training Program	MEME	Metathesaurus Editing and Maintenance Environment
ITER	International Toxicity Estimates for Risk	MEO	Medical Education and Outreach
ITK	Insight Toolkit	MEPI	Medical Education Partnership Initiative (Africa)
ITP	Informatics Training Program	MeSH	Medical Subject Headings
ITS	Internal Transcribed Space	MHC	Major Histocompatibility Complex
ITSC	Information Technology Service Center	MHL	Medical Heritage Library
ITSMS	Information Technology Service Management System	MID	Manuscript Identifiers
IUPAC	International Union of Pure and Applied Chemistry	MICAD	Molecular Imaging and Contract Database
JDBC	Java Database Connectivity	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
		MITP	Medical Informatics Training Program

<u>Acronym</u>	<u>Meaning of Acronym</u>	<u>Acronym</u>	<u>Meaning of Acronym</u>
MLA	Medical Library Association	NICHHD	National Institute of Child Health and Human Development
MLAA	Medical Library Assistance Act	NICHSR	National Information Center on Health Services Research and Health Care Technology
MLB	Medical Language Branch (database server)	NIDCD	National Institute on Deafness and other Communication Disorders
MLP	Molecular Libraries Program (at NIH)	NIDCR	National Institute of Dental and Cranio-facial Research
MMDB	Molecular Modeling DataBase	NIDDK	National Institute of Diabetes, Digestive, and Kidney Diseases
MMS	MEDLARS Management Section	NIEHS	National Institute of Environmental Health Sciences
MMTx	MetMap Technology Transfer	NIGMS	National Institute of General Medical Sciences
MOHW	Ministry of Health and Social Welfare (Tanzania)	NIH	National Institutes of Health
MOR	Medical Ontology Research	NIHMS	NIH Manuscript Submission
MOU	Memorandum of Understanding	NIH PI	NIH Pathways to Independence Award
MPCI	Missing Persons Community of Interest	NIMH	National Institute of Mental Health
mRNA	Messenger Ribonucleic Acid	NIMHD	National Institute on Minority Health and Health Disparities
MS	Mass Spectrometry	NINDS	National Institute of Neurological Disorders and Stroke
MTHSPL	Metathesaurus Structured Product Labels	NIOSH	National Institute for Occupational Safety and Health
MTI	Medical Text Indexer	NIS	Network Information Service
MTIFL	Medical Text Indexer First Line	NIST	National Institute of Standards and Technology
MTMS	MeSH Translation Management System	NLM	National Library of Medicine
MUSC	Medical University of South Carolina	NLM LitArch	NLM Literature Archive
NAC	Network Access Control	NLP	National Language Processing System
NA-MIC	National Alliance of Medical Image Computing	NMAETC	National Minority AIDS Education and Training Center
NAML	Network of African Medical Libraries	NN/LM	National Network of Libraries of Medicine
NAS	National Academy of Sciences	NNMC	National Naval Medical Center
NASA	National Aeronautics and Space Administration	NNO	National Network Office
NCATS	National Center for Advancing Translational Sciences	NOAA	National Oceanic and Atmospheric Administration
NCBC	National Center for Biomedical Computing	NOSC	Network Operations and Security Center
NCBI	National Center for Biotechnology Information	NOVA	National Online Volumetric Archive
NCCS	NIHI Consolidated Collocation Site	NPL	National Priorities List (as for Superfund)
NCHS	National Center for Health Statistics	NQF	National Quality Forum
NCI	National Cancer Institute	NRCBL	National Reference Center for Bioethics Literature
NCRR	National Center for Research Resources	NSAT	NIH Security and Accreditation Tool
NCVHS	National Committee on Vital and Health Statistics	NSF	National Science Foundation
NDC	National Data Codes	NTCC	National Online Training Center and Clearinghouse
NDF-RT	National Drug File – Reference Terminology	OA	Open Access
NEC	Necrotizing Intravascular Coagulation	OACF	Onsite Alternate Computing Facility
N _e HC	National e-Health Collaborative	OAM	Office of Administrative Management
NEI	National Eye Institute	OARF	Outreach Activity Reporting System
NewSTEPS	Newborn Screening Technical assistance and Evaluation Program	OCCS	Office of Computers and Communications Systems
NGI	Next Generation Internet	OCHD	Coordinating Committee on Outreach, Consumer Health and Health Disparities
NHANES	National Health and Nutrition Examination Surveys	OCIO	Office of the Chief Information Officer (NIH)
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		

<u>Acronym</u>	<u>Meaning of Acronym</u>	<u>Acronym</u>	<u>Meaning of Acronym</u>
OCPL	Office of Communication and Public Liaison	PHMSA	Pipeline and Hazardous Materials Safety Administration (PMA)
OCR	Optical Character Recognition	PHP	Public Health Partners
OCT	Office of Clinical Toxicology (SIS)	PHR	Personal Health Record
OD	Office of the Director	PHS	Public Health Service
ODIMRC	Office of the Disaster Information Management Research Center	PI	Pathway to Independence award
ODS	Office of Dietary Supplements (NIH)	PI	Principal Investigator
OERC	Outreach Evaluation Resource Center	PIC0	Personally Identifiable Information Comparison, and Outcome
OHA	Office of Health Affairs (DHS)	PID	Pathway Interaction Database (NCI)
ORF	Original Release Format	PIN	Precise Ingredient (term type), RxNorm
OHIPD	Office of Health Information Programs Development	PIV	Personal Identify Verification
OMB	Office of Management and Budget	PL	Person Locator
OMIA	Online Inheritance in Animals (database)	PLAWARe	Programmable Layered Architecture With Artistic Rendering
OMIM	Online Mendelian Inheritance in Man (database)	PMC	PubMed Central
OMSSA	Open Mass Spectrometry Search Algorithm	PMCI	PubMed Central International
ONC	Office of National Coordinator (for Health Information Technology)	PMC ID	PubMed Central Identification (number)
OPA	Outreach Project Application	PNG	Portable Network Graphics (file format)
OPASI	Office of Portfolio Analysis and Strategic Initiatives	PRS	Protocol Registration System
OPD	OLDMEDLINE Serials Application	PSD	Public Services Division
OPD	Outreach Products Database	PubMedHh	PubMed for Handhelds
ORDR	Office of Rare Diseases and Research (NIH)	PUE	Power Use Efficiency
OSA	Optical Society of America	PUG	PubChem Power User Gateway
OTB	Orfeo Toolbox	QA	Quality Assurance
ORWH	Office of Research on Women's Health	QCIM	Quarterly Cumulative Index Medicus
OSIRIS	Open Source Independent Review and Interpretation System	QVR	Query, View and Report
OSPB	Outreach and Special Populations Branch (SIS)	RAC	Real Application Clusters
OSTP	Office of Science and Technology Policy (White House)	RACCE	Residents' Assessment of Competency in Colposcopy Exam
PA	Pharmacological action	RAM	Random Access Memory
PAHO	Pan American Health Organization	RCDC	Research Condition and Disease Categorization
PAL	Potential Abstract Labels (tool)	RCSB	Research Collaboratory for Structural Bioinformatics
PBM	Pharmacy Benefit Manager	RDMS	Rare Disease Maintenance System
PCA	Personal Computer Advisory Committee	RefSeq	Reference Sequence (database)
PCM	Preservation and Collection Management Section (LO)	RefSNP	Reference SNP (database)
PCOR	Patient Centered Outcomes Research	RELACIGER	Red Latinoamericana de Centros de Información en Gestión del Riesgo de Desastres (Latin American Network of Risk Management Centers)
PCR	Polymerase Chain Reaction	REMM	Radiation Even Medical Management
PDA	Personal Digital Assistant	RF2	Release Format version 2
PDRS	Publisher Data Review System	RFA	Request for Applications
PDB	Protein Data Bank	RFID	Radio Frequency Identification
PDF	Portable Document Format	RFP	Request for Proposals
PDL	Personal Digital Library	RIDeM	Repository for Informed Decision Making
PDM	Patient Data Management	RML	Regional Medical Library
PDQ	Physician Data Query	RNA	Ribonucleic Acid
PFIF	Person Finder Interchange Format	RNAi	RNA Interference
PheGenI	Phenotype-Genotype Integrator	RPS-BLAST	Reversed Position Specific BLAST
PHLIP	Public Health Law Information Project	RQS	Request Submission and Tracking System
PHII	Public Health Informatics Institute	RSILC	Rotation- and Scale-Variant Line-based Color-aware (descriptor)
		RSS	Really Simple Syndication

<u>Acronym</u>	<u>Meaning of Acronym</u>	<u>Acronym</u>	<u>Meaning of Acronym</u>
RTECS	Registry of Toxic Effects of Chemical Substances	TDI	3D Informatics (Group)
RTLS	Real Time Location System	TEHIP	Toxicology and Environmental Health Information Program
RVDS	Remote Virtual Dialogue System	TERA	Toxicology Excellence for Risk Assessment
RWD	Responsive Web Design	TIC	Trusted Internet Connection
RWJF	Robert Wood Johnson Foundation	TICAP	Trusted Internet Connection Access Partners
SA&A	System Assessment and Authorization	TIE	Telemedicine Information Exchange
SAB	Source Abbreviations	TIFF	Tagged Image File Format
SBIR	Small Business Innovation Research	TKMT	Traditional Korean Medical Terms
SCID	Severe Combined Immunodeficiency	TILE	Text to Image Linking Engine
SCR	(MeSH) Supplemental Chemical Records	TIOP	Toxicology Information Outreach Project
SDK	Software Development Kit	TOXLINE	Toxicology Information Online
SEF	Serials Extract File	TOXNET	Toxicology Data Network
SEIM	Security Event and Information Management System	TPA	Third Party Annotation (database)
SEO	Search Engine Optimization	TMS	Track Management System (NCBI)
SEP	Special Emphasis Panel	TREF	Terminology Representation and Exchange Format
SEPs	Special Emphasis Panels	TRI	The Toxics Release Inventory
SEQ	Structured Evidence Queries	TS	Terminology Service
SIDA	Swedish International Development Agency	TSA	Transcriptome Shotgun Assembly
SIG	Special Interest Group	TSD	Technical Services Division
SII	Scalable Information Infrastructure	TT	Teaching Tool
SIRS	SNOMED CT International Request Submission System	TTP	Turning the Pages
SIS	Specialized Information Services	UC	Unified Communications
SITK	Simple Insight Toolkit	UCUM	Unified Codes for Units of Measure
SKR	Semantic Knowledge Representation	UID	Unique Identifier (PubMed)
SMART	Scalable Medical Alert and Response Technology	UIMA	Unstructured Information Management Architecture
SNOMEDCT	Systematized Nomenclature of Medicine Clinical Terms	UKPMC	United Kingdom PubMed Central
SO	Signing Official	UMLS	Unified Medical Language System
SOAP	Simple Object Oriented Protocol (formerly Simple Object Access Protocol)	UMLSKS	UMLS Knowledge Source Server
SOC	(HHS) Secretary's Operation Center	UN	United Nations
SPER	System for the Preservation of Electronic Resources	UNII	Unique Ingredient Identifiers Project (MeSH)
SPIN	Shared Pathology Informatics Network	UPS	Uninterrupted Power Supply
SPIRS	Spine Pathology Image Retrieval System	USHIK	United States Health Information Knowledgebase (team)
SPL	Structured Product Labels (FDA)	UTS	UMLS Terminology Services
SPWG	Special Projects Workgroup (NIH)	UUD	User Uploaded Data
SRA	Short Read Archive	VAST	Vector Alignment Search Tool
SRS	Substance Registration System (FDA)	VDB	Vertical Database
STB	Systems Technology Branch	VHA	Veterans Health Administration
STEM	Science, Technology, Engineering and Math	VOIP	Voice over IP
STR	Short Tandem Repeat	VSAC	Value Set Authority Center
STTP	Short-Term Trainee Program	VTT	Visual Tagging Tool
STTR	Small Business Technology Transfer Research	WADEM	World Congress on Disaster and Emergency Medicine
STS	Sequence Tagged Site	WGS	Whole Genome Shotgun
SVM	Support Vector Machine	WHR	Women's Health Resources
TA	Title Abbreviation(s)	YEP	Year End Processing
TBL	The bottom line		

*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

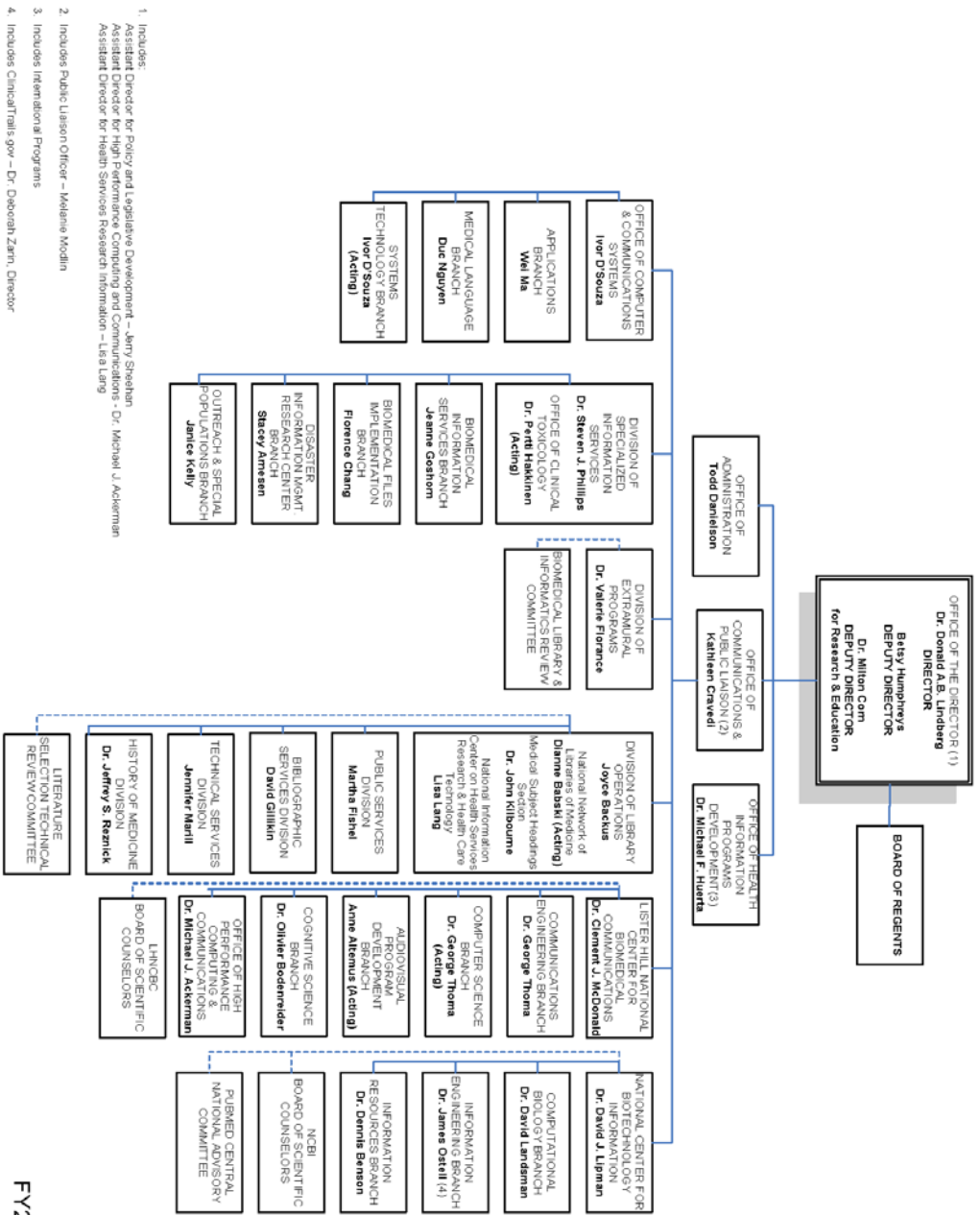
E-mail: publicinfo@nlm.nih.gov

Web: www.nlm.nih.gov

Cover:

The National Library of Medicine's National Center for Biotechnology Information (NCBI) marked its 25th anniversary in 2014. The cover displays graphically the 25 years of NCBI's data and user services' growth (1989-2014).

National Library of Medicine



1. Includes:
 Assistant Director for Policy and Legislative Development – Jerry Sheehan
 Assistant Director for High Performance Computing and Communications - Dr. Michael J Ackerman
 Assistant Director for Health Services Research Information – Lisa Lang
2. Includes Public Liaison Officer – Melaine Modlin
3. Includes Informational Programs
4. Includes Clinical Trials.gov – Dr. Deborah Zarn, Director

----- Dotted lines indicate connectors to advisory committees.

FY2014

2014

NATIONAL LIBRARY OF MEDICINE

PROGRAMS + SERVICES

 U.S. National Library of Medicine

FOOD
PATHOGENS
PROJECT

MEDGEN
PUBREADER

GENETIC
TESTING
REGISTRY
CLINVAR

1000
GENOMES

GENOME
REFERENCE
CONSORTIUM

GENOME
WIDE
ASSOCIATION
STUDIES

NIH PUBLIC
ACCESS

PUBCHEM

PUBMED
CENTRAL

HUMAN
GENOME


PUBMED

OMIM

GENOMES

GENBANK
AT NCBI

ENTREZ
BLAST


NCBI
NATIONAL
CENTER FOR
BIOTECHNOLOGY
INFORMATION

@
25

NATIONAL LIBRARY OF MEDICINE • 2014 • PROGRAMS AND SERVICES

NIH PUBLICATION NO. 14-256

1989 • 1990 • 1991 • 1992 • 1993 • 1994 • 1995 • 1996 • 1997 • 1998 • 1999 • 2000 • 2001 • 2002 • 2003 • 2004 • 2005 • 2006 • 2007 • 2008 • 2009 • 2010 • 2011 • 2012 • 2013 • 2014