THE LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

A research division of the U.S. National Library of Medicine

TECHNICAL REPORT LHNCBC-TR-2009-003

The Lister Hill National Center for Biomedical Communications Annual Report FY2009

Clement J. McDonald, M.D.
Director

U.S. National Library of Medicine, LHNCBC
8600 Rockville Pike, Building 38A
Bethesda, MD 20894

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS FY2009 ANNUAL REPORT

Clement J. McDonald, MD
Director

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

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

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

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

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

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

NLM Personal Health Record

The goal of the NLM Personal Health Record (PHR) project is to help individuals manage the care of their elderly parent(s) and/or young children and themselves. The PHR also serves as a test bed for validating and improving NLM clinical vocabularies, for studying consumers' use of PHR systems, for studying the potential of PHR-based educational reminder systems to improve prevention, and as a possible vehicle for gathering information from patients during clinical trials.

As reported last year, the NLM PHR supports the entry and tracking of key measurements and test results, prescriptions, problems and immunizations. It will produce digital and paper copies of its contents in various formats, and users can get access to MedlinePlus information resources about any prescription drug and medical condition or surgery, that they enter into the system by clicking on an icon adjacent to this the name of this drug, condition or surgery It also attaches codes to the medications, observations, and problems it carries. These codes come from terminologies supported by NLM and designated as national standards by HHS. The automatic inclusion of these codes within the NLM PHR enables both the reminders about preventative care and the automatic downloading of clinical information to the PHR.

The LHNCBC is now working with NLM and others to ensure that policies concerning the PHR are developed and in place before we deploy this system at NIH. However, outside organizations have expressed an interest in deploying the software themselves. A nearby hospital will be the likely first user.

In the last year, we have continued to expand and improve the capabilities of the PHR. We have expanded browser support beyond the original Mozilla browser to Internet Explorer 8, Opera, and Apple Safari, and we are finishing a suite that runs all software tests on each of these browsers running on each of the major operating systems. We have also finished the machinery needed for recovering user IDs and passwords via secret questions and have implemented a friendly mechanism for identifying incomplete or erroneous data entry when the user attempts to store the record.

We are finishing a major revision of the decision rules authoring system. In the first version, authors had to use a syntax that was like a programming language and required programming skill. In our new version, such skill is not needed. Instead, users pick from a series of menus and input fields on a web form. The effort is quite similar to what is required to author a Microsoft Access query. To build a base rule, the author identifies the PHR table of interest, e.g. medications, conditions, allergies, etc., and then identifies the subject and the predicate of the rule. A rule author can identify a subject by name, or class. For example, if the users picked medications as their table of interest, they could identify the subject by name e.g. Simvasatin oral, or by class, e.g. cholesterol lowering drug, and then add criteria such as status = active, or date started was before January 1, 2009. We could define another rule to determine whether a patient's last LDL cholesterol was greater than 130 mg/dl during the last year. These two rules could be combined into a third rule for reminding PHR record holders to ask their doctor about the use of diet and/or a cholesterol lowering agent if they their LDL cholesterol was elevated, *and* their medication list did *not* contain any active cholesterol lowering agents. Of course, what we have described here is for illustrative purposes only and real rules will have to be more complicated.

We have added a number of other capabilities to the PHR including:

1. a special capability for generating pop-up forms for capturing information from almost any questionnaire or laboratory panel. The only prerequisite is that the information that defines the questionnaire (or lab panel) must be in the PHR's master knowledge database and follow the LOINC form structure. Currently the PHR has access to over 200 such "questionnaires."
2. special controls to manage user corrections and deletions of information they entered about conditions, prescriptions, or other content during a previous data entry session. For example, to change a previously-entered medication record, the users would have to click on the medication of interest and then choose the desired action, e.g. discontinue, revise, or delete.
3. new tables for identifying medication content by ingredient and a system based on this table to warn PHR users when they enter a new prescription with the same ingredient as previously entered one.

This young project addresses the longstanding NLM interest of facilitating health care management and is part of the NLM strategic plan. It will help tune the message and vocabulary standards that NLM has supported and also provide another consumer entry point to a rich trove of patient-oriented data. Early research projects will focus on user needs, usability, and usage patterns to guide the next round of development and research.

Use of Surescripts prescription data in direct patient care

Surescripts operates the largest electronic prescribing network in the U.S. It handles over 2 billion prescriptions per year and can provide prescription benefit and history information for 65 percent of all patients nationally. Funded by the Bethesda Hospitals Emergency Preparedness Partnership, a collaboration formed by three Bethesda area healthcare facilities (National Naval Medical Center, NIH Clinical Center, and Suburban Hospital), this project studies the feasibility and value of using prescription history information from Surescripts in direct patient care. This external source of medication information that can be obtained automatically could be both time and life saving in disaster circumstances, when the normal route of obtaining patients' medication history is likely to be disrupted due to over-strained medical staff or special patient circumstances (e.g. unconscious patients).

As part of this project, the developers built a system that connects the Emergency Department (ED) of Suburban Hospital to Surescripts data center. The system (1) receives ADT messages generated by the HL7

engine of Suburban Hospital, (2) filters in five essential key information about the patient (first and last names, gender, date of birth, and zip code), (3) sends them to Surescripts, (4) receives prescription history of the patient if the patient's record is found in Surescripts databases, and (5) maps the information to the patient's record in the project database. All messages that are sent and received are in HL7 format. The system was installed at the data center of Suburban Hospital in the first quarter of 2009 and has been in operation since then.

In order to study the value of this additional source of information, Surescripts data was collected in parallel with the usual manual medication history for 3 months, during which the Surescripts information was not shared with the medical staff (or integrated to the information flow of the ED for patient care). The study collected information on 10,000 Emergency Department patients. Surescripts was able to match 61 percent of these patients in their database and returned some current prescription information for 45 percent. The conventional medical history identified 67 percent of patients on home medications. 39 percent of patients had drug information from both sources. 6 percent of patients had drug information only from Surescripts. Preliminary analysis suggests Surescripts information could significantly improve the completeness of the medical history data.

After the first three month of the study period, the Surescripts data was made routinely available to the medical staff. After the triage nurse enter the demographic information of the patient into the hospital system, a one-year-long prescription history of the patient with all necessary details is automatically printed in the ED in a matter of seconds.

The current challenge is to find the best way of presenting this data in a comprehensive, easily understandable but not overly-detailed format. Seamless integration of this additional source of information with the existing workflow is the key to its acceptability and utilization which will ultimately improve patient care.

In the current information flow plan, the raw Surescripts data is reprocessed, clustered by medication names, sorted by date, associated with dispensed drug duration information (when available), and at the end, a one-year-prescription history of the patient is plotted on a timeline graph.

Vocabulary Developments and Related Collaborations (e.g. with Centers for Medicare and Medicaid Service)

LHNCBC developed and released RxTerms, a drug interface terminology derived from RxNorm to support efficient and user-friendly prescription data entry in e-prescribing systems, and medication histories within electronic medical records and personal health records. The LHNCBC continues to collaborate with the Centers for Medicare and Medicaid Services (CMS) to provide guidance, content and vocabulary lists (for drugs and for common problems) to help them implement their Continuity Assessment Record Evaluation (CARE) data entry form. CMS is using RxTerms in their demonstration project for CARE. LHNCBC is using this same vocabulary for drugs and problem in the PHR we are developing.

We anticipate getting anonymyzed data about the drugs and problems entered into the CARE form from CMS when they complete their pilot project in the early part of 2010. That data will provide a powerful estimate of the adequacy of the entries and the synonymy of RxTerms and the medical problems we provided them.

Standards for Identifying Clinical Observations, Forms and Panels

In FY2009, LHNCBC expanded the LOINC database. In collaboration with NLM, HRSA, CMS, and HHS, LHC refined and moved a code set and HL7 implementation guide for clinical genetic test reporting (LOINC codes for the variable and SNOMED CT codes for the answers) through HL7 ballot approval.

Research continues on the development of an approach to the reporting of clinical genetics studies. This collaborative effort with HTSP, Partners of Boston, Intermountain Healthcare, HL7, and NCBI resulted in the creation of 12 LOINC panels and 100 new LOINC terms and a balloted HL7 implementation guide that describes how to report the results of such studies.

De-identification Tools

De-identification is a process that enables research on clinical documents by alleviating ethical concerns about protecting personal health information and privacy. The purpose of the project is to produce high-quality software to convert clinical documents into a redacted form that is compliant with the Privacy Rule of the Health Insurance and Accountability Act of 1996. The provisions of the rule dictate removal of 18 individually identifiable health information elements that could be used to identify the individual, the individual's relatives, employers, or household members.

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

We obtained narrative reports for more than 70 thousand individuals (under Institutional Review Board exemptions) to improved de-identification. The record mix includes radiology reports, history and physical exam records, occupational therapy notes, discharge summaries, referrals, consult notes, laboratory data and nursing notes. These records are being annotated with the assistance of a software tool that helps the annotators to tag all HIPAA identifiers in the reports. The tool has reduced the annotation time several fold and helped us to reach our goal of 10,000 fully annotated documents.

The de-identification software system operates through a sequence of pipelined processes: (1) Health Level 7 (HL7) message parsing, (2) part of speech tagging, (3) protected personal health information identification, and (4) redaction. The algorithmic approach aims to replicate the gold standard. The preliminary results suggest that the current system performance is measured as greater than 95 percent and close to 100 percent in both sensitivity and specificity. Current efforts are directed toward conserving all clinically important information while ensuring none of the individually identifiable health information slips through the final redacted version of the narrative.

EMR Database Research and Development

As part of our medical record research and development, the LHNCBC developed a general purpose longitudinal database structure and populated it with de-identified electronic medical record data from the MIMIC II data base under a restricted use MOU. This most recent version of this EMR data set carries clinical data for over 30,000 ICU encounters and more than 25,000 patients. It includes 11,626,411 laboratory results and over 400,000 radiology reports and discharge summaries. Altogether, it carries nearly 200 million discrete observations. It now includes data about death within the hospital, many sensitive measures of clinical status including Glasgow coma score and arterial–alveolar difference, vital signs, and enough other information to calculate a reasonable Apache score on the day of admission.

We are now testing the feasibility of using this data base for clinical research by exploring published relationships between survival and clinical variables such as the relationship between glucose control and survival. We can almost duplicate the results of the early randomized trial and at least one retrospective trial showing the benefits of tight glucose control. Our ultimate goal is to develop general tools for exploring hypotheses via formal statistical methods applied to de-identified longitudinal data bases. The MIMIC EMR data set provides the complexity and size needed to test and tune the database design and explore machinery for hypothesis generation and testing.

Newborn Screening Coding and Terminology Guide

Newborn screening is an important part of public health, but use of test results is complicated by wide variations among states in the ways tests are conducted and results recorded and by inefficient, paper-based communications. The current situation can delay rapid attention to a child's health problems, and it creates frustration and extra work for parents, health care providers, and public health authorities. During FY2009, the LHNCBC designed, created, beta tested and deployed a new Web site, the Newborn Screening Coding and Terminology Guide, as a translator to promote more efficient electronic exchange of newborn screening information.

The Web site is designed to help states move toward the use of common terminology and coding and message standards. The ultimate goal is the delivery of newborn screening results to electronic health records and public health database. Developers worked with several other agencies to organize standard definitions and codes

(using NLM standards) for more than 100 newborn screening conditions and the test measurements used to detect them. For all the conditions and tests, the Web site indicates the preferred standard terminologies and codes. Users of the Web site can view the information interactively or download electronic datasets of standard names and identifiers for use in their systems. The Web site also includes draft guidance to help newborn screening laboratories create an HL7 version 2.x message using these codes, and an annotated example HL7 message. Links to Genetics Home Reference and other resources provide additional information about newborn screening.

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

Biomedical Imaging and Multimedia

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

Imaging Tools for Biomedical Research

In FY2009, the Center developed and deployed initial versions of imaging tools to enable research in cervical cancer. In collaboration with oncologists and gynecologists at the National Cancer Institute, the team developed a suite of tools: the *Boundary Marking Tool, Multimedia Database Tool, and Virtual Microscope.* Using these tools, our collaborators are conducting research in: developing new biopsy protocols for pre-cervical cancer screening; visual evaluation of the cervix for pre-cancer detection; analyze expert inter-observer agreement; evaluating non-expert health workers to screen for cryotherapy treatment of HPV-positive women; risk factors for Inflammatory Breast Cancer; Phase II clinical trial of treatment of Kaposi sarcoma by nicotine patch; compare efficacy of glass slide diagnosis versus *virtual microscope* diagnosis for lung cancer. Researchers also developed an *HPV Linear Array tool* for the genotyping of HPV types, an important facet of ongoing cancer research work. Our NCI collaborators and their associates are using these tools in the U.S. and in several countries abroad.

As part of the tool building effort, we have also developed the *Teaching Tool*, an image-based knowledge assessment system, for administering, grading and reporting on examinations for medical professionals. The tool is a Web-browser based application using state-of-the-art software technology, including a PHP interface to MySQL and Postgres databases on the server, and a client that uses AJAX and Javascript for efficient response to the user. The American Society for Colposcopy and Cervical Pathology (ASCCP) is using the Teaching Tool to assess the professional knowledge of colposcopists and experts in cervical cancer. Developers created database content for two specific exams of the ASCCP, developed program logic to handle their particular requirements, and created a publicly-accessible version of the Teaching Tool to support two beta tests. These extensive beta tests promise the routine and nationwide use of the tool by the ASCCP. In September 2009, the Board of Scientific Counselors favorably reviewed all the tools developed in this project.

Interactive Publications Research (IPR)

This project demonstrates a type of highly interactive multimedia document that serves as a model for next-generation publishing in biomedicine. The project focuses on the standards, formats, authoring and reading tools necessary for the creation and use of such interactive publications (IP) containing media objects relevant to the biomedical literature: text, video, audio, bitmapped images, interactive tables and graphs, and clinical DICOM images such as x-rays, CT, MRI, and ultrasound.

In FY2009, researchers developed the Panorama tool for viewing and analyzing video, DICOM clinical images, tables, graphs and animations. Panorama, written in Java, was one of 9 semi-finalists out of 70 entrants in Elsevier's Grand Challenge contest.

Ongoing work is in developing the Java-based Forge as an authoring tool for IP. Forge allows authors to develop interactive articles using a set of wizards, while permitting them to write content in their original word processor; developing the capability to allow open community support to add viewers for new media types (e.g., electrocardiograms and other waveform datasets) in an open software plug-in framework; and enhancing Panorama to support a virtual file system layer to 'hide' the physical file storage location and format of a self-contained IP instance. This last feature will eliminate the need for time-consuming FTP downloads before readers can view and use a large IP.

Developers from the NLM Office of Computer and Communications Systems (OCCS) are working on another interactive publications project with the Optical Society of America (OSA). The goal is to produce four interactive issues of regularly published OSA journals and to provide a free interactive publication reader similar to the free Adobe PDF Reader. To date, OSA has published two Medline indexed interactive publications on the web. These first two articles contain 24 papers and 287 linked available interactive datasets. The next two planned issues are under development and planned for publication by February 2010. A one-year assessment of readers, focusing on the added value versus the work it takes to have access to the value, is now in progress. The preliminary results are very encouraging.

Interactive Science Publication

Data associated with medical research and clinical trials is often sequestered and not available to the greater scientific community for analysis and review. Moreover, convenient tools for viewing and analyzing data are not generally available, source data is often only indirectly linked to journal articles, and digital data is published in non-standard formats with limited ability to search metadata characterizing the associated measurements. NLM is partnering with the Optical Society of America to develop an interactive software and database infrastructure that enables viewing and analysis of curated, supplemental biomedical source data published in conjunction with peer-reviewed manuscripts; to evaluate the educational value of such an infrastructure; and to explore the problems of archiving this medium. In order to accomplish these goals, OSA is publishing four electronic special issues of OSA journals on research topics which lend themselves to interactive publishing. Articles published in these special issues are peer reviewed and fully citable as OSA journal publications indexed in Medline. They will be published on the web in Acrobat format (PDF) with links to source data, videos, and other media objects. The links allow users to quickly and conveniently download these objects and visualize them using interactive viewing software designed to look like an Acrobat plug-in. The viewing software is freely available as a download for all computer platforms. The journal articles and data sets are open access and the source data and associated metadata are searchable and accessible from outside the publication. The publications are being followed by evaluations of the usability and educational value of this form of publication. The project will also serve as an NLM test bed for archiving this new form of publication.

The first special issue, a survey issue on various biomedical topics of interest involving optics, was published in October 2008, http://www.opticsinfobase.org/oe/virtual_issue.cfm?vid=68. The second special issue focusing on *Optical Coherence Tomography in Ophthalmology* was published in March 2009, http://www.opticsinfobase.org/oe/virtual_issue.cfm?vid=81.http://www.opticsinfobase.org/oe/??????????. The third special issue focusing on *Digital Holography and 3-D Imaging* was published in October 2009, http://www.opticsinfobase.org/ao/virtual_issue.cfm?vid=94. The fourth special issue focusing on *Image Cancer Detection* will be published in March 2010.

The project includes three independent evaluations: a usability evaluation of the interactive interface; a controlled evaluation of the value of the interactivity; and a user evaluation of the perceived value. Dr. Ben Shneiderman of the University of Maryland has completed the evaluation of the user interface. Dr. James L. Mulshine of the Rush University Medical Center is conducting the controlled evaluation. Foresee, Inc. of Ann Arbor, MI with guidance from Dr. Fred Wood and Dr. Elliot Siegel of NLM, is conducting the user evaluation over the Web.

Computational Photography Project for Pill Identification

In a national effort to promote patient safety, OHPCC plans to create an authoritative, comprehensive, public digital image inventory of the nation's commercial prescription solid dose medications. Heretofore, we have conducted most of this work in partnership with the NLM Specialized Information Systems Division and the U.S. Veterans Administration to study content-based retrieval methods for medical image databases. Researchers have developed computer vision approaches for the automatic segmentation, measurement, and analysis of solid-dose medications. In particular, recent focus has been on robust color classification tools to help identify prescription drugs. In FY2009, researchers started a new project, *Computational Photography Project for Pill Identification*, funded under the American Reinvestment and Recovery Act to create a national collection of digital photographs of prescription tablets and capsules, creating high resolution digital photographs of the front (obverse) and back (reverse) surfaces of pharmacy samples, confirming that the images match the description of the medication, developing and matching the images of the samples to relevant metadata (including size descriptions, dimensions, color, and the provenance of the sample). We believe that there are between 5,000 and 32,000 prescription solid-dose oral medications to be cataloged through this process.

Multimedia Database R&D

The Virtual Microscope (VM) and Virtual Slides (VS) are an archive of virtual slides developed from the teaching set of glass slides from the Department of Pathology of the Uniformed Services University and other collaborating institutions. An entire slide is digitized, segmented and processed to simulate an examination of a glass slide under the microscope but with a Web browser. The collection preserves the specimen for posterity and allows viewing by users worldwide anytime. Recent additions include annotations and automatic linking to Medline/PubMed. A related collection of images from the Armed Forces Institute of Pathology (AFIP) fascicles allows users to search images and automatically link to Medline citations. Viewing of VS with mobile devices is a recent application. Collaboration with the Massachusetts General Hospital explores its use in distance education and training. We are developing other collaborations to study its use in telemedicine and teleconsultation.

Virtual Microscope (VM) and Virtual Slides

An archive of virtual slides has been developed from the teaching set of glass slides from the Department of Pathology of the Uniformed Services University and other collaborating institutions. An entire slide is digitized, segmented and processed to simulate an examination of a glass slide under the microscope but with a Web browser. The collection preserves the specimen for posterity and allows viewing by users worldwide anytime. Annotations and automatic linking to Medline/PubMed has been enabled. A related collection of images from the AFIP fascicles allows users to search images and automatically link to Medline citations. Viewing of VS with mobile devices is a recent application. A recent collaboration with the Massachusetts General Hospital explores its use in distance education and training. Other collaborations will study its use in telemedicine and teleconsultation.

Content Based Image Retrieval (CBIR)

Our research in Content-Based Image Retrieval has several objectives. One is to develop techniques to augment our existing cancer research tools with automated processes. For example, the CervigramFinder automatically indexes and allows retrieval of cervigrams (and spine x-rays) using shape, color and texture features. This system contains the key elements needed to augment the Boundary Marking Tool with an automated assist for the user in marking boundaries of regions of medical significance. The Cervigram Segmentation Tool combines several image segmentation and shape similarity algorithms for research into selecting the optimal algorithms that enable automated segmentation for image indexing. MOSES is a tool provided as a service for evaluation of segmentations by multiple experts, to reduce the inter- and intra-observer variability in marking boundaries of significant regions in cervigrams.

In addition, a hybrid system (SPIRS/IRMA) has been developed by linking geographically separated CBIR systems to exploit the characteristics of each. SPIRS/IRMA allows retrieving detailed medical image data in large collections of heterogeneous images of varying modalities, presentations and anatomy. For example, a user may narrow a search to *spinal x-rays* (using IRMA at University of Aachen) and then retrieve specific x-rays containing *osteophytes* at the SPIRS system at NLM.

CBIR is also a component of a process for indexing illustrations in medical journals by using image features, figure captions and in-document text mentions. This research is aimed at enriching the user experience of searching for relevant documents by including the contents of medical images, photographs, graphs and other illustrations found in articles. Techniques developed in this work were evaluated in the international ImageCLEF competition and found to be successful.

Another avenue explored in this research area is distributed computing and use of GPUs for compute-intensive CBIR tasks, with a particular focus on image segmentation.

The Visible Human Project

The Visible Human Project image data sets are designed to serve as a common reference for the study of human anatomy, as a set of common public domain data for testing medical imaging algorithms, and as a test bed and model for the construction of image libraries that can be accessed through networks. The Visible Human data sets are available through a free license agreement with the NLM. They are distributed to licensees over the Internet at no cost; and on DAT tape for a duplication fee. More than 3,000 licensees in 58 countries are applying the data sets to a wide range of educational, diagnostic, treatment planning, virtual reality, and virtual surgeries, in addition to artistic, mathematical, legal, and industrial uses. The Visible Human Project has been featured in more than 900 newspaper articles, news and science magazines, and radio and television programs worldwide.

FY2009 saw the continued maintenance of two databases to record information about Visible Human Project use. The first, to log information about the license holders and record statements of their intended use of the images; and the second, to record information about the products the licensees are providing NLM in compliance with the Visible Human Dataset License Agreement.

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

In FY2010 the LHNCBC will hold a cross-disciplinary workshop made up of key representatives of the disciplines that are relevant to bone size and shape in an effort to find a source for the needed statistical data. These disciplines will include forensic pathology and other forensic sciences, paleontology, radiology, orthopedics, anatomists, prosthesis designers, clothing designers and furniture designers. By inviting the participation of the interested community combined with using Web tools developed under the Interactive Publication Project, developers expect to collect the needed data.

3D Informatics

In FY2009, the 3D Informatics Program (TDI) expanded research efforts concerning problems encountered in the world of three-dimensional and higher-dimensional, time-varying imaging.

The LHNCBC provides continuing support for image databases, including ongoing support for the National Online Volumetric Archive (NOVA), an archive of volume image data. This collection contains 3D data from across medicine. Contributors to the collection include the Mayo Clinic Biomedical Imaging Resource and the Walter Reed Army Medical Center Radiology Department. The archive contains such integrated and multimodal data as virtual colonoscopy matched with recorded video from endoscopic interventions, time-varying 3D cardiac motion, and 4D MRI of a human hand. During 2009, the TDI group installed the necessary software and hardware infrastructure, including a Linux server and a MIDAS software system, to support interactive scientific publication at NLM.

Beyond data collection, we continue to analyze and develop software tools for studying biological problems such as tumor micro-environments and the transmission of HIV AIDS (from macrophage to T-cell) through 3D optical and electron microscopy. This work combines high performance computing with life sciences research, accelerating and empowering investigators in the detection and prevention of cancer and infectious diseases. Members of the TDI group contributed to research in HIV virus identification in dual-beam electron microscopy of dendritic and macrophage cells. Other work included geometric analysis of mitochondrion in dual-beam data taken of metastatic cancer cells in the human liver. This work was conducted in support of the National Cancer Institute Laboratory for Cell Biology.

Collaborative work in telemedicine has increased the demand for large display technologies, but the cognitive pathways for understanding how the human visual system processes large-scale digital displays are not well understood. Throughout 2009, researchers continued to develop high-end rendering systems for large displays, incorporating multiple GPUs in volume-rendering applications capable of rendering the full VHP Male dataset at real time rates. We began investigations in the question of presence and visualization, combining forces with the NCRR-funded Biotechnology Resource Center at the Utah Scientific Computing and Imaging (SCI) Institute and their access to the Magnetic Encephalography (MEG) equipment in the University of Utah Department of Radiology. We also began a study of the difference between human-scale versus non-human scale perception, seeking changes in cortical processing of reading and eye-tracking from large wall displays versus hand-held book-sized displays. The result of such a study could have a far-reaching impact on how new digital user interfaces are developed in future generations.

The Center also began a study of the use of rapid prototyping technologies, to develop witness objects or phantoms, reference models as a public engineering and scientific standard for research in image-based computer-aided diagnosis. We have characterized the x-ray attenuation characteristics of some of the 3D-printing materials available at NIH and are presently evaluating the use of contrast agents as printing materials to vary the appearance of the 3D models when viewed with x-ray CT scanners. The goal is to create complex, anatomically-accurate models to test diagnostics systems and evaluate and compare their performance under known conditions. This work is conducted in partnership with the National Institute of Allergy and Infectious Diseases.

Insight Tool Kit (ITK)

The Insight Toolkit is approaching its tenth year of development with an official software release of ITK 3.16 in September 2009. Over 845,000 lines of openly available source code comprise ITK, making available a variety of image processing algorithms for computing segmentation and registration of high dimensional medical data on a variety of hardware platforms. ITK can be run on Windows, Macintosh, and Linux platforms, reaching across a broad scientific community that spans over 40 countries and more than 1500 active subscribers to the global software list-serve. A consortium of university and commercial groups, including OHPCC intramural research staff, provide support, development, and maintenance of the software.

ITK is an essential part of the software infrastructure of many projects across and beyond the NIH. The Harvard-led National Alliance of Medical Image Computing (NA-MIC), an NIH Roadmap National Center for Biomedical Computing (NCBC), has adopted ITK and its software engineering practices as part of its engineering infrastructure. ITK also serves as the software foundation for the Image Guided Surgery Toolkit (IGSTK), a research and development program sponsored by the NIH National Institute for Biomedical Imaging and Bioengineering (NIBIB) and executed by Georgetown University's Imaging Science and Information Systems (ISIS) Center. IGSTK is pioneering an open API for integrating robotics, image-guidance, image analysis, and surgical intervention. International software packages that incorporate ITK include *Osirix*, an open-source diagnostic radiological image viewing system available from a research partnership between UCLA and the University of Geneva and the Orfeo Toolbox (OTB) from the Centre Nationale D'Etudes Spatiales, the French National Space Administration.

From 2002 to 2008, the ITK effort has attempted to stimulate scientific development through supplementing research platforms such as the Analyze from the Mayo Clinic, SCIRun from the University of Utah's Scientific Computing and Imaging Institute, and the development of a new release of VolView, free software for medical volume image viewing and analysis. In 2009, OHPCC and the ITK project began participation in American Reinvestment and Recovery Act economic stimulus efforts. OHPCC is developing a new acquisition plan for a software revision and development program to upgrade ITK for emerging computational platforms for the upcoming decade. The development of a major new version of ITK will help to continue our international leadership role in medical imaging research.

Image and Text Indexing for Clinical Decision Support and Education

As part of the Clinical Information Systems effort, the Image and Text Indexing project seeks to automatically identify illustrations in biomedical articles that could provide multi-media assistance to clinical decision making. We developed an experimental search engine that illustrates retrieval results with images extracted from scientific

articles. The retrieval methods implemented in the search engine achieved top performance in the ImageCLEF 2009 case-based retrieval task. This task reflects real-life clinical situations in which a clinician is looking for publications related to a specific patient's case. The quality of our retrieval results in this task allows for evaluation in a teaching hospital that is planned for the next year.

Automated mapping of text from a small teaching collection of electrocardiogram images to UMLS concepts was evaluated for both accuracy and utility in indexing the image. Although a substantial proportion of the mappings were judged as matches, a minority were judged as useful for indexing the image. Most images were judged to need additional descriptors beyond those found in the accompanying text. The additional descriptors suggested were also highly likely to map to an existing UMLS concept or combination of concepts. We propose that sample medical image collections automatically indexed and reviewed by experts might provide a template for more precise automated indexing. The manuscript has been submitted to JAMIA.

Observations in published clinical case reports were matched by hand to entries in the Logical Observations Identifiers Names and Codes (LOINC) in a preliminary analysis of the feasibility of automated natural language processing. Coverage of history, physical, laboratory, and imaging findings was nearly complete, and categorization of terms assigned suggests a strategy for automatic coding. This study was presented as a poster (Proc AMIA Symp, 2009, 1036).

Researchers plan to further study the influence of augmenting bibliographic references retrieved from a database search with images; new ways to combine text and image retrieval results; new ways of organizing and presenting retrieval results using annotated images; and further improvement in the automatic single and multi-panel image extraction, annotation, and complementary text extraction.

InfoBot

As part of the Clinical Information Systems effort, the **InfoBot** project enables a clinical institution to automatically augment a patient's emergency medical record (EMR) with pertinent information from NLM and other resources. The InfoBot software runs as background agents, both at the institution and at NLM. APIs would be supplied to the institution to allow them to integrate the search setup and to display and store results in their existing EMR system.

In FY2009, we developed two prototypes and made them available to collaborators from 4 departments at the NIH Clinical Center (CC). The NLM InfoBot server has handled over 2000 CC requests since the system went live.

The first prototype, a table driven, asynchronous InfoBot system, was evaluated by clinicians at the Clinical Center, and results presented at the 16th National Evidence-Based Practice Conference. This was followed by the development of a second (Web-based) prototype capable of real-time processing of clinical text, generation of database and search-engine queries, summarization and display of the result. This prototype was integrated with CRIS, and made available through the Evidence-Based Practice tab in CRIS July 2009. In FY2010, we plan to complete the evaluation of this prototype at the Clinical Center, implement a module for automatic creation of rulesets by users, replace/expand single-document summarization of retrieval results with multi-document summarization, improve automatic processing of clinical notes, and improve automatic query generation.

Turning The Pages

This project allows users to turn and view page images in a photorealistic manner on touch-sensitive monitors in kiosks, as well as 'click and turn' in an online version. The goal of the project is to provide the lay public a compelling experience of historically significant and normally inaccessible books in medicine and the life sciences.

In FY2009, we released the kiosk version of the Edwin Smith Papyrus, having first created a 3D model for flat scrolls. In addition, we released both kiosk and online versions of Brunschwig's Liber de Arte Distillandi, printed in Strasbourg in 1512, a practical manual on chemical, alchemical, and distillation devices and techniques used to manufacture drug therapies. For offsite demonstrations, we developed standalone projector versions of all eight TTP books. New work includes the conversion of Hanaoka, a Japanese medical manuscript, to TTP format.

Biomedical Image Transmission via Advanced Networks (BITA)

Research was conducted in the performance of NLM-supported networks such as the Interactive Video Outreach and Distance Learning Network for Minority High School Students, the BHEPP network, 802.11a, 802.11b and higher designation wireless network implementations, and networks exhibiting narrow bandwidths, high latency and high jitter.

We evaluated image quality, usability, and speed performance for multiple image sources, including large image file transfers and high definition videoconferencing over Internet2, through formal demonstrations in collaboration with universities and research centers nationally and internationally.

As part of this effort, research staff represented NLM and NIH at an Internet2 Advisory Council, Joint Engineering Team, Large Scale Networking Team, CSIA (Cybersecurity and Information Assurance) and other forums for high performance/speed networks.

AUTOMATED CONCEPT EXTRACTION FROM DOCUMENTS

Research in this area is directed toward developing techniques and algorithms to extract bibliographic data from biomedical journal articles, both digitized and Web documents, to build MEDLINE citations. The projects in this category are MARS and its various spin-offs and the Indexing Initiative. These systems address the NLM Goal 1: *Seamless, Uninterrupted Access to Expanding Collections of Biomedical Data, Medical Knowledge, and Health Information*.

Medical Article Records System (MARS)

The goal of this project is to develop and operate systems to efficiently extract bibliographic data from scanned and Web journal articles to populate MEDLINE. For paper-based journals, the system delivers bibliographic data by combining document scanning, optical character recognition (OCR), and rule-based and machine learning algorithms. The MARS production system currently in operation extracts and organizes bibliographic data by using advanced algorithms for automated zoning, field identification, and syntax reformatting. In addition, biomedical lexicons are used to implement pattern matching algorithms to correct errors in OCR-detected affiliation information, and reducing incorrectly highlighted words for increased operator productivity.

In FY2009, all engineering support for the offsite MARS production facility was provided: installation of upgraded modules, testing, maintenance and operation of all hardware and software for servers, clients and networks, and the necessary system administration. A significant addition was the incorporation of Unicode in the Edit and Reconcile modules that enable operators to view and enter special characters, diacritics and foreign language text in a *wysiwyg* manner (exactly as seen in the published article.)

As a key element in helping to achieve goals of NLM's Indexing 2015 Initiative, the Publisher Data Review (PDR) system was developed and currently undergoing field testing. This system is designed to provide operators data missing from the XML citations sent in directly by publishers such as databank accession numbers and NIH grant numbers, thereby lowering the manual effort in completing citations for MEDLINE. In addition, incorrect data sent in by the publishers can be corrected by PDR. The current manual effort in filling in missing data and correcting wrong data from the publishers is considerable since the operators generally have to look through an entire article to find this information, and then key them in. Currently, we are designing a machine learning method for extracting Investigator Names, a labor-intensive effort if done manually since there could be hundreds of such names listed in an article.

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

ACORN

The ACORN initiative (*Automatically Creating OldMedline Records for NLM*) aims to capture bibliographic records from pre-1960 printed indexes (e.g., IM, QCIM, QCICL, etc.) for inclusion in NLM's OldMedline database, thereby creating a complete record of citations to the biomedical literature since Index Medicus appeared in the late 19th century. In FY 2009 we continued our investigation of scanning, image enhancement, OCR, image analysis,

pattern matching, and related techniques to extract unique records from the printed indexes. Researchers designed a prototype consisting of three main components: Quality Control, Processing, and Reconcile. The Quality Control module is completed, and work is proceeding toward implementation of the Processing module which will group OCRed text into records for operators to verify using the Reconcile component. In FY2010, a prototype will be delivered to Library Operations as a pilot system.

MARG

As part of the AIDE project, a validated test set for document image analysis has been created for the computer science and informatics communities for research into advanced algorithms for data mining. This Medical Article Records Groundtruth (MARG) database has attracted more than 17,000 visits from 99 countries. It consists of images of journal articles, the corresponding OCR data, zones, labels and verified data obtained from the routine operation of the MARS production system.

Indexing Initiative

The Indexing Initiative (II) project investigates language-based and machine learning methods for the automatic selection of subject headings for use in both semi-automated and fully automated indexing environments at NLM. Its major goal is to facilitate the retrieval of biomedical information from textual databases such as MEDLINE. Team members have developed an indexing system, Medical Text Indexer (MTI), based on two fundamental indexing methodologies. The first of these calls on the MetaMap program to map citation text to concepts in the UMLS Metathesaurus which are then restricted to MeSH headings. The second approach, a variant of the PubMed related articles algorithm, statistically locates previously indexed MEDLINE articles that are textually related to the input and then recommends MeSH headings used to index those related articles. Results from the two basic methods are combined into a ranked list of recommended indexing terms, incorporating aspects of MEDLINE indexing policy in the process. Image Indexing Initiative (I3) has used interactive MetaMap successfully for automated mapping of terms suggested by subject matter experts to Metathesaurus concepts. I3 will continue efforts to improve the precision and recall of these mappings with additional features of the MetaMap API.

The MTI system is in regular, increasing use by NLM indexers to index MEDLINE. MTI recommendations are available to them as an additional resource through the Data Creation and Maintenance System (DCMS). This year MTI recommendations are being augmented by the attachment of subheadings to some of the MeSH headings it recommends. Indexers now have the option of accepting MTI heading/subheading pairs in addition to unadorned headings. In addition, indexing terms automatically produced by stricter version of MTI are being used as keywords to access collections of meeting abstracts via the NLM Gateway. These collections include abstracts in the areas of AIDS/HIV, health sciences research, and space life sciences.

Indexing Initiative development focuses on testing and updating recently added functionality to the MTI system such as the inclusion of subheading attachment recommendations and the addition of an explanation facility to inform indexers how MTI arrived at specific MeSH recommendations. In addition a recent effort consisted of successfully modifying MTI for NLM Cataloging; the cataloging version of MTI is now in regular use. System-related objectives for II include completing the final testing of the migration of our systems to a Linux environment.

Digital Preservation Research (DPR)

This project addresses an important problem for libraries and archives, viz., to retain electronic files for posterity, both documents in multiple formats (e.g., TIFF, PDF, HTML) as well as video and audio resources. Researchers focus on the preservation of digitized documents, and have built and deployed a first version of a prototype System for Preservation of Electronic Resources (SPER). SPER builds on open source systems while incorporating inhouse-developed modules that implement key preservation functions: ingesting, automated metadata extraction and file migration. NLM curators have started using SPER to preserve more than 60,000 court documents from the historic medico-legal FDA collection. In FY2009, approximately 10,000 documents were processed and made Web-accessible to the public.

Research continues into techniques for digital preservation with the goal of improving SPER while processing the remaining FDA documents, as well as investigating ways to accommodate other NLM collections:

(a) Improve accuracy of the automated metadata extraction (AME) system by enhancing the layout classification model and the pattern recognition module. (b) Examine other Support Vector Machine implementations for use in SPER. (c) Enhance the extraction model to encrypt metadata patterns and search/extraction rules for each document layout for the FDA collection. (d) Improve the pattern search engine to identify and extract metadata from a broader set of historic documents using this model.

INFORMATION RESOURCE DELIVERY FOR CARE PROVIDERS AND THE PUBLIC

The Lister Hill Center performs extensive research in developing advanced computer technologies to facilitate the access, storage, and retrieval of biomedical information.

Clinical Research Information Systems

ClinicalTrials.gov provides the public with comprehensive information about all types of clinical research studies, both interventional and observational. At the end of FY2009, the site had nearly 80,000 protocol records, over 800 of which display summary results, sponsored by the U.S. Federal government, pharmaceutical industry, academic and international organizations from all 50 States and in 171 countries. Some 36 percent of the trials listed are open to recruitment, and the remaining 64 percent are closed to recruitment or completed. ClinicalTrials.gov receives over 40 million page views per month and hosts approximately 900,000 unique visitors per month. Data are submitted by over 6,300 study sponsors through a Web-based Protocol Registration System, which allows providers to maintain and validate information about their studies, including summary results data.

ClinicalTrials.gov was established by the National Library of Medicine (NLM) in 2000 in response to the Food and Drug Administration Modernization Act of 1997 and to support NLM's mission of disseminating biomedical knowledge and advancing public health. ClinicalTrials.gov was enhanced in FY2009 to implement additional requirements of Section 801 of the Food and Drug Administration Amendments Act of 2007 [Public Law 110-85]. The law required the expansion of the registry and the addition of a results database. Thus, the ClinicalTrials.gov registry was expanded in November 2007 to include additional information about applicable drug and device clinical trials. New links from registration records to the FDA Website (e.g., Drugs@FDA) were added and existing links to NLM's MEDLINE and DailyMed Web sites were enhanced. As a consequence of this law, new registrations in FY2009 have risen to an average of 330 per week, a 32 percent increase from FY2007, the year prior to enactment of the law (average 250 per week). In September 2008, ClinicalTrials.gov launched the "basic results" database, which complements the registry. It includes tables of summary results data on primary and secondary outcomes of registered trials, as well as information on the patient populations studied. In its first year of operation, over 1,400 results records were submitted by 300 study sponsors. The average number of submissions has increased, ending FY2009 with an average of 50 results records submitted per week. To help study sponsors with submitting data to the new results database, a number of documents were developed, including "helpful hints," a summary of common errors, and a guide to results data entry. While an optional module for reporting serious and frequent adverse events observed in a clinical trial was made available in September 2008 for public use and testing, a modified version of the module became mandatory by law on September 27, 2009. The modifications made to the adverse event module were based on feedback from stakeholders and consultation with risk communication experts. As called for by law, NIH held a public meeting on issues related to the expansion of the clinical trial registry and results database on April 20, 2009. In addition to comments presented orally at the meeting, over 60 written comments regarding issues related to the expansion have been submitted to the public docket. The expanded registration requirements as well as the results database will be further implemented through rulemaking and NLM is working with the Food and Drug Administration (FDA) on the proposed rule. ClinicalTrials.gov continues to work on other aspects of the law, including, but not limited to, expansion of the results database, a pilot quality control study, and consulting with risk communication experts. When fully implemented, the registry and results database will become a unique resource for scientific and clinical information that can assist in providing patients, healthcare providers, and researchers more comprehensive information about ongoing and completed research.

ClinicalTrials.gov was actively involved in FY2009 in educating the public and the regulated community on the new law, and continuing to promote standards of transparency in clinical research through trial registration. This information was communicated to a broad range of U.S. and international stakeholders via presentations and peer-reviewed publications. As a result of increasing awareness of the law and the importance of trial registration,

nearly 17,000 new registrations were received over FY2009. ClinicalTrials.gov continues to collaborate with other registries, professional organizations, and regulators in working towards developing global standards of trial registration and reporting to results databases.

Genetics Home Reference

Genetics Home Reference (GHR) is an online resource that offers basic information about genetic conditions and the genes and chromosomes related to those conditions. This resource provides a bridge between the public's questions about human genetics and the rich technical data that has emerged from the Human Genome Project and other genomic research. Created for the general public, particularly individuals with genetic conditions and their families, the GHR web site currently includes user-friendly summaries of more than 475 genetic conditions, more than 680 genes, all the human chromosomes, and mitochondrial DNA. The web site also includes a handbook called *Help Me Understand Genetics*, which introduces users to fundamental topics in human genetics including mutations, inheritance, genetic testing, gene therapy, and genomic research. Recent additions to the Handbook include information about the validity and reliability of genetic tests.

Genetics Home Reference celebrated its sixth anniversary in 2009. In the past year, the project expanded its genetics content for consumers. Specifically, GHR staff added more than 200 new summaries to the web site in FY2009, an increase of about 17 percent from the previous 12 months. Staff intend to continue this rate of production in FY2010, covering additional Mendelian genetic disorders as well as more complex disorders. The team also plans to continue expanding the Gene Families feature, which currently includes explanations of about 40 families of related genes.

Usage of the GHR web site continued to increase in FY2009. This year, the site averaged more than 14,100 visitors per day (an increase of about 9 percent over the previous 12 months) and more than 18.8 million hits per month (an increase of 14 percent over the previous 12 months). GHR continues to be recognized as an important health resource.

This year, GHR staff performed outreach activities to increase public awareness of the web site. The project continues to support the Information Rx initiative, a free program that enables doctors and nurses to write "prescriptions" directing patients to the GHR web site for an explanation of genetic disorders and related topics. In other outreach activities, GHR staff presented the web site to several visiting groups, including educators and students, and represented the project at meetings and conferences. Staff members will continue to educate others about this important resource in FY2010.

Profiles in Science Digital Library

The Profiles in Science Web site (Profiles) showcases digital reproductions of items selected from the personal manuscript collections of prominent biomedical researchers, medical practitioners, and those fostering science and health. Profiles in Science provides researchers, educators, and potential future scientists worldwide access to extraordinary, unique biomedical information previously accessible only to patrons able to make an in person visit to the institutions holding the physical manuscript collections. Profiles also serves as a tool to attract scientists to donate their collections to archives or repositories in order to preserve their papers for future generations. Profiles in Science decreases the need for handling the original materials by making available high quality digital surrogates of the items. Standardized, in-depth descriptions of each item make the materials widely accessible, even to individuals with disabilities. The growing Profiles in Science digital library provides ongoing opportunities for future experimentation in digitization, optical character recognition, handwriting recognition, automated image identification, item description, digital preservation, emerging standards, digital library tools, and search and retrieval. In October 2008, Profiles in Science was awarded the C. Herbert Finch Online Publication award at the Mid-Atlantic Regional Archives Conference.

The content of Profiles in Science is created in collaboration with the History of Medicine Division of NLM, which processes and stores the physical collections. Several collections have been donated to NLM and contain published and unpublished materials, including manuscripts, diaries, laboratory notebooks, correspondence, photographs, poems, drawings and audiovisual resources. The collections of Victor A. McKusick and Adrian Kantrowitz were added this year. Currently 26,347 digital items composed of 139,224 image pages are available on Profiles in Science. Presently the Web site features the archives of twenty-nine prominent individuals:

Christian B. Anfinsen
Virginia Apgar
Oswald T. Avery
Julius Axelrod
Paul Berg
Francis Crick
Rosalind Franklin
Donald S. Fredrickson
Edward D. Freis
Alan Gregg
Michael Heidelberger
Adrian Kantrowitz
C. Everett Koop
Arthur Kornberg
Mary Lasker
Joshua Lederberg
Salvador E. Luria
Barbara McClintock
Victor A. McKusick
Marshall W. Nirenberg
Linus Pauling
Martin Rodbell
Florence R. Sabin
Wilbur A. Sawyer
Maxine Singer
Fred L. Soper
Sol Spiegelman
Albert Szent-Györgyi
Harold Varmus

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

In addition to releasing new Profiles in Science collections during FY2009, LHNCBC staff performed analyses and made modifications to the Profiles in Science systems. They analyzed the Profiles in Science access log files to compare with ForeSee Results report of the number of American Customer Satisfaction Index (ACSI) surveys completed. During the survey period, 813,647 unique IP addresses accessed Profiles in Science. Of these, 233,811 unique IP addresses were offered the survey at least once. In total, 1,021 respondents completed the ACSI survey at least once, which is the equivalent of 1.25 per 1,000 unique IP addresses, or .0125 percent, that accessed Profiles during the survey period. On an average day during the survey period, 3,670 unique IP addresses accessed Profiles; 60 percent were new and 40 percent had visited previously. Developers also continued to port the Data Entry Program from MS Access Visual Basic for Applications to a Web based application, and continued to port the Solaris version of Profiles in Science to Linux. They also modified, debugged and added reports to the Diagnostic Server, a tool that assists in quality control efforts and provides a read-only view of the Profiles in Science database. Staff debugged the software that logs technical metadata about the digital items, and they rebuilt the software that exports the database to a standard text file. They continued to write software to perform tasks that were done manually or in a semi-automated fashion. Developers debugged, modified and upgraded software underlying the Profiles in Science systems as well as the LHNCBC Web site, the Clinical Questions database, and the LHNCBC Personnel System to adhere to HHS, NIH and NLM security requirements. Staff evaluated software to extract text from PDF files and for making video clips adhere to Section 508 accessibility requirements.

Nursing Home Screener (NHS)

NHS is a Web2.0 system using MySQL, Google Maps, and data from CMS to quickly locate and select nursing homes. A user can search for homes by zip code or town, and receive a list of homes in order of overall quality (denoted by a one to five star rating), as well as a map showing locations. This list may be filtered by such characteristics as non-profit ownership or accepting Medicare or Medicaid payments. Activities in FY 2009 included: developing a beta Web site, with substantially revised design based on 2008 usability testing; making data updates more automatic, reducing ambiguous location data; conducting further usability testing; reporting the project at AMIA '08 and the Board of Scientific Counselors; and receiving an AMIA Distinguished Paper Award.

FY2010 plans include: launch a public site; implement initial, minimal customized page for each nursing home with tailored information; enhance user interface based on usability testing; run monthly data freshening; add statistics and graphics; pair customized page with per-home blog, and seed blog with news clippings about nursing homes.

Evidence Based Medicine - PubMed for Handhelds

PubMed for Handhelds was publicly released in FY2003. Developed to facilitate evidenced-based medical practice with Medline access at the point of care via smartphones, wireless PDA's, netbooks or portable laptops, PubMed for Handhelds requires no proprietary software and reformats the screen display as appropriate for the wireless handheld device being used. In support of evidence-based clinical practice, clinical filters feature easy access to relevant clinical literature. Newly developed resources allow searching Medline through text-messaging. An algorithm to derive “the bottom line” (TBL) of published was recently added for a clinician’s quick reading at the point of need.

New features can create a "consensus" opinion of multiple publications. Recent collaborative projects are ongoing to extend its reach to Botswana, Africa and the Pacific Islands.

User Focused Portals: NLM Gateway

The NLM Gateway is an ongoing production system that provides results from 24 NLM information resources in response to a single query. Since these resources are frequently updated, improved, and otherwise modified, the Gateway must change with them. Periodic changes to the NLM Document Type Definition (DTD), to MeSH and the MeSH mapping file, and to the UMLS Metathesaurus are accommodated each year. More than 110,000 meeting abstracts are indexed using the tools of the Indexing Initiative. Access to more than 70,000 Images from the History of Medicine was added this year.

Ongoing usability studies have resulted in further improvement of several aspects of the user interface. A user survey sought to determine the degree of user interest in a mobile version of the NLM Gateway for systems like the iPhone.

COMMUNICATION INFRASTRUCTURE RESEARCH AND TOOLS

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

Videoconferencing and Collaboration

Staff continued to investigate, review, and develop collaboration tools, research their application, and use the tools to support ongoing programs at the NLM.

Work on uncompressed high definition video over IP continued. A second iHDTV node with streams of 1.5 gbps was installed and experiments were also done with UltraGrid uncompressed HD video having similar bit rates. In addition, a compressed version of UltraGrid with bit rates of 250 mbps was tried out successfully with the Rochester Institute of Technology. HD video was tested using new network protocols for dynamic network switching and an upgrade of the OHPCC/Collab subnet to 10 gbps was planned to accommodate HD bandwidth. A demonstration was done at the Internet2 Spring Member Meeting. Two versions of a pan, tilt, zoom (PTZ) remote camera control program were developed for Sony HD cameras (a standalone version and one for use on the AccessGrid) in anticipation of using HD video in telemedicine research. Work continues on additional HD enhancements for telemedicine research. The camera control program was made peer to peer for AccessGrid, while a peering standalone program is under development. Work on a more Windows native version of the University of Washington's iHDTV software and a version allowing bandwidth adjustments have been suspended because ConferenceXP, an open source program originally from Microsoft that was spun off the University of Washington also has HD capability. An uncompressed version was tested locally on the same machine and a compressed version with bit rates from 1 to 5 mbps was tested externally. The plan is to test iHDTV and ConferenceXP over optical connections between machines connected back to back in preparation for external tests when the network upgrade is realized. Staff is working on a manuscript comparing HD and other network video technologies used in the Collab.

Staff continue to look for partners for testing uncompressed HD video in dermatology, since the Puget Sound VA, which help formulate the clinical study, could not participate because it became committed to implementing a regional store and forward teledermatology program. Staff have approached the dermatology departments at the Washington Hospital Center and Medical University of South Carolina. Both have expressed initial interest. Staff will continue exploring other clinical areas where HD research may be appropriate.

A manuscript describing results of clinical research study of video medical interpretation completed at the Medical University of South Carolina is undergoing revision for publication. The study compared conventional standard definition video to phone and in person interpretation. A follow up study using lower quality video (less than full screen) and cell phone technology has just started. Extensive tests were done with VSee, a low bandwidth video program, in anticipation of the study.

The results of an earlier co-location (presence) study of videoconferencing with the University of Alabama at Birmingham has been accepted for publication. A follow up study using one way video and chat to study learning

outcomes and collaborative behaviors when students were co-located and dispersed was conducted, but the results still need to be analyzed.

Work continued with SIS on distance education outreach program for minority high school students and an article was published on expanding the program to Alaska. Farrington High School in Hawaii has been added to the program. The school has a diverse Native Hawaiian, Native Pacific, and Asian student body. It joins Kotzebue High School and King Drew Medical Magnet High School in Los Angeles, with their respective Native Alaskan and African American and Hispanic student populations. Programs continued to be archived and placed on the Collab web and streaming servers.

An alliance was developed with the NIH Library to continue to offer NCBI database training using the Virtual Computer Lab methodology developed earlier in cooperation with the University of Puerto Rico (UPR). An article describing the Virtual Computer Lab methodology was published and this year programs were conducted with the University of North Carolina at Chapel Hill and the Charles R. Drew University of Medicine and Science. Additional quarterly programs are being planned with UNC and Duke.

PTZ program and other collaboration programs developed by OHPCC continue to be put on the Collab web server for others to download as they are refined. Documentation for the SIS distance learning program was added to the iHDTV, UltraGrid, and Virtual Computer Lab documentation put online last year.

Staff continued a retrospective review of NLM OHPCC Telemedicine, Next Generation Internet, and Scalable Information Infrastructure Projects with the objective of identifying computing, communication, and health science application themes in this diverse body of work. A draft manuscript was developed and is currently being edited for submission.

OHPCC Collaboratory for High Performance Computing and Communication

The “Collab” was originally established as a resource for researching, testing, and demonstrating imaging, collaboration, communications and networking technologies related to NLM’s Next Generation Network initiatives. This infrastructure is now used by staff to keep abreast of and test new technologies of possible interest to NLM (and others in biomedical informatics) and to conduct ongoing imaging, collaboration and distance learning research within OHPCC. The technology infrastructure is used to collaborate with researchers outside the NLM and, when appropriate, it is leveraged to support other activities and programs of the NLM. The facility can be configured to support a range of technologies, including 3D interactive imaging (with stereoscopic projection), the use of haptics for surgical planning and distance education, and interactive imaging and communications protocols applicable to telemedicine and distance education involving a range of interactive video and applications sharing tools. The latter enable staff to collaborate with others at a distance and, at the same time, demonstrates much of the internal and external work being done as part of NLM’s Visible Human and advanced networking initiatives. The collaboration technologies include a complement of tools built around the H.323 and MPEG standards for transmitting video over IP and open source technologies such as the Access Grid.

BabelMeSH

BabelMeSH is a multilanguage and cross-language search tool for healthcare personnel who prefer to search MEDLINE/PubMed in their native languages. Journals’ language of publications can be selected. Through international collaborations, including WHO Eastern Mediterranean Regional Office in Cairo, users can now search in Arabic, Chinese, Dutch, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish, Swedish and English. Some specialty organizations are using BabelMeSH as a tool to search their collection of images. Through Google, immediate translation of MEDLINE abstracts to French, Italian, Portuguese and Spanish has been added recently.

PICO Linguist

PICO (Patient, Intervention, Comparison, and Outcome) Linguist is an application available through BabelMeSH that allows users to search Medline/PubMed in a more clinical and evidence-based manner. This work is significant because it is the only cross-language search portal on the Internet that allows the input in more than two languages. It is also unique because it allows the user to search in character-based languages (non-Latin alphabet), transform it

to an English language search, and retrieve citations published in any language or language combination. Full-text articles may be linked to the result available online without subscription requirements.

Computing Resources Projects

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

The network management team plans, implements, tests, deploys, and operates high-speed network connectivity locally as well as over Internet and Internet-2. The core projects include studying and planning the implementation of central network management for effectively responding to network alerts and malfunctions; 10 Giga BPS network to support research projects that require high-speed communication capacity; and an enterprise device management system to update large number of network devices uniformly.

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

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

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

DocView Project: Tools for Using and Exchanging Library Information

The goal of this project is to conduct R&D on advanced tools allowing libraries and users to access biomedical information. In FY 2009, research focused on the development of MyDelivery, a new Internet communications system designed to deliver very large files and large numbers of files, especially over potentially unreliable networks such as wireless used by an increasingly mobile population. Health science applications often require the use and exchange of information contained in very large files (e.g., digitized x-ray images, sonographic images, digital video files, MRI, CT scans, PET scans, and scanned document images). Targeted for use in clinical, research, administration, and library environments, the MyDelivery system will be capable of reliably communicating biomedical information contained in files of any size over networks of all types, including potentially unreliable ones.

In FY2009, following alpha testing, we conducted an extensive beta test with 200 outside testers. Ongoing work involves completing the MyDelivery API, modifying the system to be fully compliant with FIPS 140-2, integrating it with the Interactive Publications Panorama viewer to demonstrate the capability of the API, and releasing MyDelivery as open source code.

- DocMorph

As part of the DocView project, research and system engineering continues to maintain and improve the operation of DocMorph, a Web-based server providing users remote image and information processing capabilities via the Internet. This system now accepts more than fifty file formats, including black and white images, grayscale and color images, text and word processing files, to produce four outputs: PDF files, TIFF files, text, and synthesized speech. DocMorph averages 1,000 conversions daily, and 1,000 unique users monthly. It is used by several hundred libraries, including NLM which uses it mainly in our interlibrary loan service.

- MyMorph

While DocMorph is generally accessed via a Web browser, the **MyMorph** client software allows users to perform large scale conversion of thousands of files at a time. MyMorph has more than 12,000 registered users, many of whom are document delivery librarians in small libraries around the country, using MyMorph as part of their daily document delivery operation.

Image Storage and Transmission Optimization (ISTO)

Investigators researched the best approaches for using advanced compression techniques, such as Wavelet transform, for the storage of digitized spine x-rays and uterine cervix images, in addition to implementing an efficient decompression algorithm. Developers incorporated wavelet decompression and multiscale display into a dissemination system for the digitized biomedical images within the WebMIRS system and the Multimedia Database Tool.

Work under way includes completing the development of the HVSQ codec incorporating Wavelet transform techniques coupled with scalar and vector quantization to achieve maximum compression consistent with image quality for the digitized x-ray images of the lumbar and cervical spine and uterine cervix images.

Other ongoing work includes making these advanced compression techniques compatible with open standards, such as Internet Imaging Protocol (IIP), and developing a standard interface for classification of vertebrae from NHANES x-ray images, and organizing feature vectors for image retrieval.

LANGUAGE AND KNOWLEDGE PROCESSING

Terminology Research and Services

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

Objectives for FY2010 are:

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

Medical Ontology Research (MOR)

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

During FY2009, interoperability issues were investigated among several pairs of biomedical ontologies, including BioTop and the UMLS Semantic Network, SNOMED CT and MedDRA, the Mammalian Phenotype Ontology and OMIM, and between RxNorm and NDF-RT.

RxNav, the standalone browser for RxNorm, NLM's drug terminology integration database, was extended in order to provide better support to our users. The Application Programming Interface (API) has been used for mapping large amounts of NDC codes to RxNorm identifiers. Feasibility of integrating other drug information sources with RxNorm through RxNav was demonstrated.

We kept investigating the benefits of using Semantic Web technologies including RDF - the Resource Description Framework - and triple stores (e.g, Virtuoso) for the integration of biomedical information and in support of translational research activities (e.g., comparing genotype-phenotype associations across species).

This year, our research activities resulted in 6 journal articles, 6 papers in conference proceedings, and 13 invited presentations. We continue to collaborate with leading ontology and terminology centers, including the National Center for Biomedical Ontology, the International Health Terminology Standards Development Organization (SNOMED CT) and the World Health Organization (ICD 11).

Clinical and Translation Science

Using a newly developed database of translational terms, RxNorm and MeSH, the LHNCBC developed a novel search tool to search for innovative, novel and promising translational research. The query is initiated using disease processes and/or interventions or both as search terms. Publications identified as translational in nature are then retrieved with relevant terms highlighted for easy recognition. With interventions, such as drugs in the RxNorm database and disease processes in MeSH that are pertinent clearly identified, the user can quickly find publications to facilitate research, experiment planning and bench-to-bedside applications. An additional benefit from this project is the inclusion of a new term, "translational research", to the MeSH vocabulary suggested by one of the project team members.

Semantic Knowledge Representation (SKR)

The Semantic Knowledge Representation project conducts basic research in natural language processing (NLP) based on the UMLS knowledge sources. The project focuses on developing SemRep to extract semantic predications from text to support innovative information management applications in biomedicine, including advanced tools for clinical decision support, practice guideline development, and literature-based discovery. Current efforts are adapting this technology to identify selected concepts in clinical narrative.

SKR researchers are developing a SemRep predication database which holds 22 million predications from 6.6 million MEDLINE citations (1999 through October, 2009). The project team is exploiting this database to find publications that support critical questions used during the creation of clinical practice guidelines (with support from NHLBI). Further, the team is collaborating with academic researchers in using the predication database to help interpret the results of microarray experiments, to support literature-based discovery, and to investigate advanced statistical methods for enhanced information retrieval.

UMLS AND CLINICAL VOCABULARY STANDARDS

This program encompasses multiple projects. The problem list vocabularies project has produced the CORE (Clinical Observations Recording and Coding) subset just released as a UMLS knowledge source. The RxTerms project is an efficient drug interface terminology that links directly to RxNorm. Inter-terminology Mapping activities with IHTSDO (International Health Terminology Standards Development Organization) and direct UMLS-related activities continue. The TREF (Terminology Representation and Exchange Format) specification will be used by NCHS (National Center for Health Statistics) to produce a TREF version of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM).

The CORE Problem List Subset of SNOMED CT

Problem lists in Electronic Health Records (EHR) facilitate the organization and communication of clinical data and reasoning. They also drive additional functions like clinical decision support, generation of reimbursement codes, auditing and research. At present, most institutions use their own problem list vocabularies for problem list entry. The creation and maintenance of these vocabularies are resource and time consuming. The lack of a shared list of

terms impairs data interoperability. The UMLS-CORE (short for Clinical Observations Recording and Encoding) Project was started two years ago to study the use of controlled vocabularies in clinical documentation at a summary level, such as in a problem list. This project has two goals:

- To study the problem list vocabularies of large health care institutions and to characterize them in terms of their size, pattern of usage, map-ability to standard terminologies and the extent of overlap
- To identify a CORE subset of terms that have a high frequency of occurrence in problem lists

Problem list terms and their frequency of usage are collected from seven large healthcare providers: Beth Israel Deaconess Medical Center, Kaiser Permanente, Mayo Clinic, Intermountain Healthcare, Regenstrief Institute, Nebraska University Medical Center and the Hong Kong Hospital Authority. The most commonly used terms, covering 95 percent of the usage volume in each institution, are mapped to the UMLS. About 8 percent of these terms cannot be mapped. The overlap among the mapped terms is assessed. The average pairwise overlap among institutions is about 40 percent. The terms that are shared among institutions are used much more frequently than terms that are not shared. Using this data, a subset of 5,000 SNOMED CT concepts, called the CORE Problem List Subset of SNOMED CT is identified and made available to the public. SNOMED CT is the U.S. Consolidated Health Informatics (CHI) designated terminology standard for problem lists. The CORE Subset represents the SNOMED CT concepts that are most frequently used in problem lists. It can be used as a starter set for the creation of problem list vocabularies. This will facilitate the use of SNOMED CT in EHRs and enhance data interoperability. The CORE Subset was first published in July 2009 and is updated regularly to synchronize with updates in SNOMED CT and the UMLS. It can be downloaded as a standalone file or extracted from the UMLS as a content view subset. Terms that are frequently found in problem lists but not present in SNOMED CT have been submitted to the International Health Terminology Standards Development Organization (IHTSDO) for consideration of addition.

RxTerms

RxTerms fills the need for a free, user-friendly and efficient drug interface terminology that links directly to RxNorm, the national terminology standard for clinical drugs. Efficiency of data entry is enhanced by systematic rearrangement of RxNorm drug names and pruning of drugs that are not likely to be useful in prescribing e.g. drugs unavailable in the U.S., allergenic extracts. Addition of common synonyms and abbreviations further improves data entry efficiency. Other usability enhancing features include ‘tall man lettering’ of look-alike drug names as recommended by the FDA, and user-friendly concentration units for liquid doses. RxTerms has been evaluated for coverage of commonly prescribed drugs and data entry efficiency. Coverage is over 99 percent for both branded and generic drugs. Data entry efficiency is improved when RxTerms is used compared to the original RxNorm drug names.

RxTerms has been available for free download since November 2008. It is updated every month to synchronize with the monthly full release of RxNorm. There are about 250 registered users to date. RxTerms is currently being used in one of CMS’s applications in the post-acute care environment. It is also used in NLM’s Personal Health Record. The use of RxTerms will facilitate the inclusion of RxNorm identifiers in Electronic Health Records and promote data interoperability.

Disaster Information Management

Lost Person Finder (LPF)

The Lost Person Finder (LPF) is a system for family reunification during a mass casualty event. It combines image capture, database and Web technologies.

In FY2009, project activities included requirements gathering, analysis, reporting, test, and demonstration. Developers built a demo version of the LPF site, customizing and enhancing the Sahana open-source disaster management system by focusing on its Missing Person Registry. Customization included a tailored LPF theme and

logo, Americanization of words and phrases in the database and user interface, and incorporation of standard ontologies and localized geographical features.

Other developmental activities included: developed TriagePic software for image capture of triage victims, and integrated Bluetooth camera image transmission and email-base record distribution; built first version of LPF Notification Wall (images of victims on a large display), including Web services for the application to access data from the LPF database, and iconic overlays to represent victim name and status. In October 2009, we participated in a large-scale multi-institutional demonstration (Collaborative Multi-Agency Exercise or CMAX) and demonstrated TriagePic, search capability, and the Notification Wall at collaborating institutions, the Navy and Suburban hospitals in Bethesda.

Plans are under way to: investigate privacy and security issues related to LPF use; add email and notification modules to LPF Web site; refine TriagePic interface in response to feedback at CMAX; customize LPF Web site for individual participating hospitals; develop the LPF Mobile Web site and create search capability for popular mobile platforms; and refine information shown on the Notification Wall for different audiences.

Video Production, Retrieval, and Reuse Project

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

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

A number of LHNCBC development projects require videographics, interactive multimedia development, imaging, animation, or video production as part of the overall project objectives. A major effort in this area is the improvement of rendering times for videographics, and 3D visuals and animations for DVD and other interactive multimedia productions. Extensive development work was directed toward the planning and demonstration of interactive multimedia for the 2010 NLM Exhibition “Native Concepts of Health and Illness.”

Training Opportunities

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

In FY2009, the Center provided training to 44 participants from 13 states and 5 countries. Participants worked on research projects including 3-D informatics, automated indexing, clinical information systems, consumer health information, content-based information retrieval, de-identification of medical records, document processing. preservation of electronic resources, image, text and document processing, information retrieval, interactive publication, medical ontology, medical terminology, mobile computing, natural language processing, personal health record, question/answering and telemedicine research.

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

The Informatics Training Program sponsors a Clinical Informatics Postdoctoral Fellowship Program to attract young physicians to NIH to pursue research in informatics. This program is run jointly by the Lister Hill Center and the Clinical Center to bring postdoctoral fellows to labs throughout NIH. Funding is from the LHNCBC. The Center continues to offer an NIH Clinical Elective in Medical Informatics for third and fourth year medical and dental students. The elective offers students the opportunity for independent research under the mentorship of expert

NIH researchers. The Center also hosts the eight-week NLM Rotation Program which provides trainees from NLM funded Medical Informatics programs with an opportunity to learn about NLM programs and current Lister Hill Center research. The rotation includes a series of lectures covering research being conducted at NLM and the opportunity for students to work closely with established scientists and meet fellows from other NLM funded programs.