

Donald A.B. Lindberg, M.D., Director
National Library of Medicine
March 6, 2007

Mr. Chairman and Members of the Committee:

I am pleased to present the President's budget request for the National Library of Medicine (NLM) for Fiscal Year 2008, a sum of \$312,562,000.

The National Library of Medicine has a remarkable track record of preserving the past while serving the present and preparing for the future. A just completed Long Range Plan done by the Library's Board of Regents lays out in broad terms the challenges the Library will face over the next decade and charts a course for action to successfully meet these challenges.

Prominent among the challenges is the need to create the information resources essential to achieving the goal of "personalized medicine," in which prevention and treatment strategies are tailored to an individual's specific genetic make-up. The first step is to provide huge linked databases and software tools that allow scientists to correlate clinical, genomic, and chemical compound data with published research findings to determine how genetics and a person's environment interact to cause disease and to identify potential new therapies. Such resources, now being developed by NLM, will speed scientific discovery and can ultimately transform medical care by allowing clinicians to customize treatments to a patient's genetic characteristics.

In an era of increasing chronic disease, a related challenge is the need to empower people with the knowledge and motivation to improve their health and play a more active role in their health care. The information that pours out of the nation's laboratories—and often finds its way into the public media—has the potential of improving the health status of our citizens. The National Library of Medicine has created heavily used Web-based information services aimed at the public. These services transmit the latest useful findings in lay language and provide guidance that can be easily understood by the public. NLM works with libraries and community-based organizations to increase public awareness and use of these valuable resources.

Electronic health records with advanced decision support capabilities will be essential to help with any effort to achieve personalized medicine and will also help people manage their own health. Much of the seminal research work in this arena was supported by the National Library of Medicine or undertaken by people who received NLM-funded informatics education. This work builds on two decades of research and development of the Unified Medical Language System (UMLS) resources which help computer systems behave as if they "understand" the language of biomedicine. The NLM also serves as an HHS coordinating center for standard clinical vocabularies and supports, develops, or licenses for US-wide use key clinical vocabularies.

No information source is useful if it is unavailable. A third major challenge facing the National Library of Medicine is ensuring uninterrupted access to critical information resources in the event of disaster or other emergency, natural or man-made. As recent hard experience demonstrated, this requires careful advanced planning,

strong inter-organizational arrangements, and skillful management of information during the emergency, in addition to robust technical backup arrangements for computer and communication systems. NLM's new Long Range Plan specifically recommends that the Library establish a new Disaster Information Management Research Center and ensure effective recognition and use of libraries as a major and largely untapped resource in the nation's disaster management efforts.

This opening statement is built around these three themes—scientific information resources that can lead to personalized medicine, information services that enable greater personal involvement in health and health care, and marshalling the Library's resources to assist the country's in emergency situations.

SCIENTIFIC INFORMATION RESOURCES—NEAR AND LONG TERM

Fueled in part by funding from the National Institutes of Health, the pace of discovery in today's world of biomedical research is amazing. The NLM is now at the center of much biomedical research—not only receiving, storing, and disseminating published research results, but actually serving as a crossroads for the genomic and other data coming from laboratories around the world. NLM databases and systems are essential tools in all aspects of biomedical research. Users conducted more than one *billion* searches of them in the last year.

The core of the National Library of Medicine is its expanding collection of more than 8 million books, journals, and other materials. The Library subscribes to more than 20,000 periodicals of which some 5,000 are indexed for Medline/PubMed, the immense online database of the journal literature. From the more than 16 million records in Medline/PubMed one may link to a tremendous variety of relevant Web-accessible online resources at NLM and elsewhere. NLM's National Center for Biotechnology Information (NCBI) has already begun building the Medline/PubMed of the future by redesigning its displays and interfaces to make it easy for users to see important links and retrieve information they might not otherwise have noticed.

The NCBI is the source of GenBank, the genetic sequence databank that contains all publicly available DNA sequences. GenBank is produced from thousands of sequence records submitted directly from researchers and institutions prior to publication. NCBI has also created PubChem, a repository for what are called “small molecules” that are crucial in drug development. Small molecules are responsible for the most basic chemical processes that are essential for life and they often play an essential role in disease.

The NCBI's effective performance on these and other trans-NIH priorities has earned NLM a prominent role in the important new Genome-Wide Association Studies (GWAS) project. GWAS is an NIH-wide initiative directed at understanding the genetic factors underlying human disease. It involves linking genotype data with phenotype information in order to identify the genetic factors that influence health, disease, and response to treatment. NCBI is building the databases to incorporate the clinical and genetic data, link them to the NLM's molecular and bibliographic resources and, for the first time, make these data available to the scientific and clinical research community. dbGaP (database of Genotype and Phenotype) debuted in December 2006 to archive and distribute data from Genome-Wide Association Studies.

PubMed Central, a Web-based archive of biomedical journal literature also developed by the NCBI for the NIH, provides free access to the full-text of peer-reviewed articles. PubMed Central is also home to full-text journal articles submitted by scientists with NIH funding under the NIH Public Access policy.

NLM's Lister Hill National Center for Biomedical Communications also produces important tools for biomedical and informatics research, including digital image libraries—sets of image data that can be used in research, clinical care, and training. In one example, NLM is currently collaborating with NIH and other researchers to develop advanced imaging analysis tools for research in human papillomavirus infection and cervical neoplasia. The tools will allow effective analysis of some 100,000 images of the uterine cervix and they will become the primary resource for professional training and testing in this field. Another set of imaging tools being widely applied in the scientific community, for education and other purposes, is related to the "Visible Humans." These two enormous data files (one male and one female) were created under the guidance of the Lister Hill Center and provide detailed image data sets that serve as a common reference for the study of human anatomy, for testing medical algorithms, and as a model for image libraries that can be accessed through networks.

INFORMATION SERVICES FOR THE PUBLIC

The audiences served by the Library have multiplied in recent years. In addition to providing researchers and health care providers with access to scientific information, the NLM also now has services for the public—from elementary school children to senior citizens. The Library's main portal for consumer health information is MedlinePlus, available in both English and Spanish. Much of this information is based on research done or sponsored by the NIH Institutes. In addition to more than 700 "health topics" (main entries on diseases and disabilities), MedlinePlus has interactive tutorials that are useful for persons with low literacy, medical dictionaries, a medical encyclopedia, directories of hospitals and providers, surgical videos that show actual operations, and links to the scientific literature.

Several databases for consumers are byproducts of research in NLM's Lister Hill Center. One of these is the ClinicalTrials.gov database, which describes clinical research studies funded by NIH and others around the world. The site contains information on more than 37,000 federally and privately supported trials and is searched daily by some 30,000 people. Another Lister Hill Center database is the Genetics Home Reference, a Web site for consumer-friendly information about genetic conditions and the genes or chromosomes related to those conditions.

NLM's toxicology and environmental health program also produces heavily used consumer information resources. The Household Products Database provides easy-to-understand data on the potential health effects of more than 2,000 ingredients contained in more than 6,000 common household products. The colorful Tox Town looks at an ordinary town and points out many harmful substances and environmental hazards that might exist there. ToxMystery, an unusual interactive Web site for children between the ages of 7–10, provides an animated, game-like interface that prompts children to find potential chemical hazards in a home.

Of inestimable help to the NLM in meeting its varied responsibilities—both to the scientific community and to the public at large—are the 5,800 member institutions of the National Network of Libraries of Medicine. The Network comprises eight Regional Medical Libraries, 120 “resource libraries” primarily at schools of the health sciences, and thousands of hospital libraries and community-based organizations. Together they form an efficient way to ensure that the published output of biomedicine is easily accessible by scientists, health professionals, and the public. They cover the critical “last mile” to familiarize researchers, health professionals and the public and to develop sustainable partnerships with community organizations to improve access to health information for underserved populations.

MANAGING VITAL INFORMATION IN TIMES OF DISASTER

A number of NLM’s advanced information services and tools are designed for use by emergency responders when disaster strikes. The Library has a history of providing assistance in such cases, for example the gas leak disaster in Bhopal, India, in the eighties, and Hurricane Mitch and the earthquakes in Central America in the nineties. NLM’s TOXNET, a cluster of databases covering toxicology, hazardous chemicals, toxic releases, environmental health and related areas, provides a foundation for services to first responders, such as WISER (Wireless Information System for Emergency Responders). Used in Louisiana after Hurricane Katrina, WISER provides information via handheld mobile devices to help identify unknown substances.

Among other such projects, the Library: (1) supported pioneering work on automated biosurveillance, self-healing wireless networks, and smart tags to track patients during emergencies; (2) built the Influenza Virus Resource with the National Institute of Allergy and Infectious Diseases to provide vaccine researchers access to genomic data of many influenza strains; (3) developed OSIRIS (Open Source Independent Review and Interpretation System), a software package to assist in identifying 9/11 victims’ remains via DNA; (4) worked via the National Network of Libraries of Medicine to re-establish and maintain a level of health information services in the Katrina-affected region; and (5) developed the Radiation Event Medical Management (REMM) system, in collaboration with the HHS Office of Public Health Emergency Preparedness, the National Cancer Institute, and the CDC.

In summary, the National Library of Medicine is well positioned to make a maximum contribution to the nation’s health—by making increasing amounts of scientific data available to researchers and health practitioners, by contributing to the national effort to improve the information infrastructure of the health care system, by providing to the public access to authoritative information for use in maintaining their personal health, and by enabling health sciences libraries to make substantial contributions of disaster information management. All of these activities will depend on a strong and diverse workforce for biomedical informatics research, systems development, and innovative service delivery. To that end, the National Library of Medicine will continue its longstanding support for post-graduate education and training of informatics researchers and health sciences librarians and redouble its efforts to improve the diversity of these fields.