DEPARTMENT OF HEALTH AND HUMAN SERVICES

NATIONAL INSTITUTES OF HEALTH

Fiscal Year 2005 Budget Request

Witness appearing before the
Senate Subcommittee on Labor-HHS-Education Appropriations

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

April 1, 2004

William Beldon, Acting Deputy Assistant Secretary, Budget

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Statement by

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

on

Fiscal Year 2005 President's Budget Request
for the National Library of Medicine

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 2005, a sum of $325,147,000, which reflects an increase of $16,671,000 over the comparable Fiscal Year 2004 appropriation.

The National Library of Medicine continues to be the premier source of science-based medical information. Just 10 years ago the Library introduced its Web site—one of the very first in the federal government—and so began a decade of amazing growth in the amount and variety of medical information it made available. Today the Library's Web service not only provides free access to Medline/PubMed, the largest and most reliable database of scientific medical information in the world, but NLM has created information products designed specifically for patients, families, and the public.

Despite its recent successes, NLM believes that the surface has barely been scratched and that the future holds the promise of many more valuable information products for the professions and the public. The Library's communications experts are at the cutting edge of new technology and, as more and more users have access to ever more powerful networks, the Library will put in place sophisticated yet easy to use information services that allow users free access to the world's burgeoning base of science-based health

information. For scientists this means access not only to the growing published journal literature, but also electronically to scientific monographs and textbooks and to a variety of genomic information resources through NLM's National Center for Biotechnology Information (NCBI). For the general public, this means making even more consumer health information—from the National Institutes of Health and other reliable sources—available from the NLM's Web site.

The new NIH Roadmap Initiative has the potential to have a profound and positive impact on how American medical research is conducted. The NLM sees itself has having an important role in the Initiative in three areas. Because the Roadmap recognizes that one of the most powerful and unifying concepts of 21st century biology is that of bioinformatics, the computerized bioinformatics databases and analysis tools of the NCBI will become even more central to the research enterprise. Second is the Roadmap's requirement to "re-engineer the national clinical research enterprise." NLM's leadership role in working with biomedical vocabularies—the Unified Medical Language System, the recently announced arrangement with the SNOMED clinical vocabulary, and NLM's expanding the NIH clinical trials database—are all key aspects of improving clinical research. Finally, the Roadmap articulates NIH's responsibility to communicate research results to improve the quality of life for all people. The Library has a central role in collecting and communicating these results through Web-based information services and online databases. These are described in what follows.

TOOLS FOR SCIENTISTS AND HEALTH PROFESSIONALS

The NLM's Medline/PubMed is the most-used database of peer-reviewed medical information in the world. It contains more than 12 million references and abstracts to the world's medical literature published since the 1960s; an ancillary "OldMedline" extends the coverage back to the early 1950s. Each year millions of scientists and health professionals connect to Medline/PubMed (no registration or fee is required) and search for information they can use in the research or practice. More than a half billion such searches are done every year. The newest system, introduced in 1997, is constantly

being improved. Several years ago NLM introduced links between Medline/PubMed references and publisher websites so users could retrieve the full text of articles. Today, more than 4,000 of the database's 4,600 publications have such links.

Another heavily used database is GenBank, the repository of all publicly available DNA sequences sent to the NLM from laboratories around the world. GenBank, and an increasing array of other valuable data resources, is the responsibility of the National Center for Biotechnology Information. The Center, which was created by the Congress in 1988 with the mandate to manage and disseminate genetic data, coordinates closely with the NIH Human Genome Project. GenBank today contains more than 27 million sequence entries totaling 33 billion base pairs from over 130,000 species. NLM, through the Web operations of the NCBI, receives more than a quarter million visitors a day seeking molecular biology information ranging from DNA sequences and protein structures to the related research literature.

A repository for chemical structure and assay data has been suggested as one aspect of NLM's involvement with the NIH Roadmap Initiative on "small molecules" to enhance research and develop new therapies. The NCBI is working on such a repository—called PubChem—which will integrate into one open database, information from existing chemical structure databases at various NIH institutes as well as data supplied from industry and academic centers. By providing chemical structure validation and structure-structure matching and by linking to descriptions of the compounds in journal articles, PubChem will play an invaluable role in making this information useful to scientists.

PubMedCentral, a digital archive, is an important component of the infrastructure needed to enhance access to the life sciences literature. Publishers electronically submit peer-reviewed research articles, essays, and editorials. NLM guarantees free access to the material; copyright remains with the publisher or the author. Access to PubMedCentral is free and unrestricted.

The full text of more than 100 life science journals, some going back decades, is now available, and more are added as they sign on to the system. Digitally archiving the scientific literature and guaranteeing access for future generations is an important NLM responsibility.

INFORMATION SERVICES FOR THE PUBLIC

The National Library of Medicine has become a favorite destination of seekers of health-related information on the Web—people looking for answers to questions about their health or the health of their loved ones. MedlinePlus, the largest of NLM's Web offerings for the general public, now receives about 4 million unique visitors a month. Increasingly, they also find their way on the NLM Web site to other services created specifically for them—NIHSeniorHealth.gov, ClinicalTrials.gov, Genetics Home Reference, Household Products Database, and Tox Town are all recent examples. These Web sites contain or point to information created by NIH components and other reliable noncommercial sources. They require NLM librarians and informationspecialists to work closely with a wide variety of outside organizations. MedlinePlus, launched in November 1998, today is one of the most heavily trafficked Web sites containing health information for the public. It has more than 650 "health topics," containing, for example, overview information, pertinent clinical trials, alternative medicine, prevention, management, therapies, the latest research, and the latest news from the print media. There are even links to the scientific literature through Medline/PubMed. In addition to the 650 health topics, there are medical dictionaries, encyclopedias, directories of hospitals and providers, and interactive "tutorials" with images and sound. MedlinePlus en español was introduced in 2002 and has grown to virtual parity with the English version. Both scored the highest marks of any Federal Web site in a recent outside evaluation. A new aspect of MedlinePlus is its plan to "Go Local," that is, to link users with community helping services near them. North Carolina is the first MedlinePlus partner to go local.

The National Library of Medicine is collaborating with the American

College of Physicians in a unique "Information Rx" project that seeks to encourage practicing physicians who are members of the College to "prescribe" MedlinePlus to their patients who need further information on a medical subject. After test runs in Georgia, Iowa, Virginia, and Florida, the Information Rx program will go nationwide later in 2004.

MedlinePlus is not the only NLM information service directed at the consumer. Another very popular resource is ClinicalTrials.gov, which integrates previously fragmented information on human studies for different conditions into a single, coherent system, providing the public with an easy-to-use and convenient "one-stop" site for comprehensive information on clinical trials. The site, which is used not only by the public but by their health care providers, currently includes information on approximately 8,800 trials for hundreds of diseases and conditions conducted in about 90 countries. ClinicalTrials.gov receives approximately 16,000 visitors daily and over 3 million page views monthly.

Late in 2003 another service for the public was launched: NIHSeniorHealth.gov. This site contains information in a format that is especially usable by seniors. For example, the site features large print and easy-to-read segments of information repeated in a variety of formats—such as open-captioned videos and short quizzes—to increase the likelihood it will be remembered. NIHSeniorHealth.gov has a "talking" function, which allows users the option of reading the text or listening to it as it is read to them.

Another new NLM consumer service is the Household Products Database. This is a guide that provides easy-to-understand information on the potential health effects of more than 2,000 ingredients contained in more than 4,000 common household products. The database provides information on many of these substances and their potential health effects, in consumer-friendly language. For more technical information, users can launch a search for a product or ingredient from the product's page into NLM's TOXNET, a cluster of databases on toxicology, hazardous chemicals, and related areas.

Another consumer health information resource introduced in 2003 is the Genetics Home Reference. Genetics is a complex subject, and much of the primary data and literature are difficult to understand without formal training. The Genetics Home Reference Website augments MedlinePlus with summaries of genetics information and an overview of the fundamentals of genetic science. The user can browse by a specific disease/condition or by gene. It also has a geographic list of genetic counselors and information for care-givers. The database has more than 100 condition summaries and 80 gene summaries and new content is being added continuously.

The Library launched Tox Town late in 2002. Tox Town looks at an ordinary town and points out many environmental hazards that might exist there. Users can click on a town location, like the school, and see a colorful dollhouse-style cutaway view of that building. Toxic chemicals that might be found in the school are listed, along with links to selected Internet resources about school environments. There are similar cutaways for offices, factories, parks, and other locations. NLM has plans to add new scenes, such as an urban community and a farming region.

SERVING SPECIAL COMMUNITIES

With all these unique information resources, it becomes more and more important for the Library to engage in outreach to let citizens know what is available. The 5,100-member National Network of Libraries of Medicine is an important partner in these outreach endeavors. Many of the programs are directed at minority populations. For example, there are programs to assist in remedying the disparity in health opportunities experienced by African Americans, Latinos, Native Americans, senior citizens, and rural populations. A new NLM database introduced in 2003 has health information aimed at Asian Americans; 2004 will see a similar database with information about the health concerns of Native Americans.

Under a program with the Historically Black Colleges and Universities (HBCUs), NLM is helping to train people to use information resources in dealing

with environmental and chemical hazards. The latest aspect of this outreach effort is NLM's collaboration with the United Negro College Fund Special Programs Corporation to work with the HBCUs in the area of consumer health to encourage the use of reliable electronic health information (such as that provided by the NLM) by the public.

NLM also has been instrumental in reaching out to other countries around the world to help improve their access to scientific medical information. The oldest such program is that involving formal partnerships with major institutions in 20 countries. The NLM helps them obtain computerized access to the literature; the countries in turn help NLM receive the medical literature from that part of the world. The Library is also a key player in the Multilateral Initiative on Malaria, the multiagency effort to improve malaria research in African nations. NLM's role is to establish and maintain the first malaria research communications network, MIMCOM. There are now 19 research sites in 9 countries participating, with full access to the Internet.

SCIENCE ADVANCES

Many scientists believe that molecular biology is the primary driver of medical advances in the 21st century. The rapidly increasing volume of molecular data and the need to decipher its cryptic and subtle patterns has created demanding requirements for computerized databases and analysis tools, special curatorial expertise, and unique physical facilities. The National Center for Biotechnology Information is a key player in ensuring that the outpouring of data from molecular biology laboratories around the world is turned to life-enhancing purposes. GenBank, as noted above, is growing rapidly with contributions received from scientists around the world. Scientists also avail themselves of sophisticated computational tools, such as the BLAST suite of programs, which lets scientists search enormous quantities of data for sequence similarities that will identify genes and genetic features. Another tool, Entrez, allows users to search DNA sequences and literature information with techniques that are fast and easy to use. The newest tool is the "Reference

Sequence Collection,'' which provides a centralized, integrated, non-redundant set of sequences that is integrated with other information for all major research organisms. Using the Reference Sequence Collection, time once spent on having to identify, gather, and analyze data can now be spent effectively on research.

The Center is now also conducting research using the human genome sequence to begin exploring the history of human populations. NCBI researchers, working with other collaborators, first assembled a set of 500,000 high-confidence variations and then compared the distribution of these variations on the genome to that predicted by several models of population history. They found that the data best fit a model in which the human population shrank dramatically about 40,000 years ago, a time when modern humans first appeared in Europe. The model suggests that the population subsequently grew about 30,000 years ago, consistent with archaeological evidence of a population expansion during that period. The results indicate that databases of genetic variation constructed alongside the human genome project can provide a unique insight into the history of human populations. This insight may also explain how these populations may respond differently to selective pressures such as infectious diseases.

NLM's Lister Hill National Center for Biomedical Communications sponsors high-technology communications research projects in such areas as high quality imagery, medical language processing, high-speed access to biomedical information, developing intelligent database systems, multimedia visualization, data mining, and machine-assisted indexing. One prominent area of research has been the Visible Human Project. The project consists of two enormous (50 gigabytes) data sets, one male and one female, of anatomical MRI, CT, and photographic cryosection images. These data sets are available through a free license agreement to 1,800 individuals and institutions in 47 countries where they are being used in a wide range of educational, diagnostic, treatment planning, virtual reality, artistic, and industrial applications. An

‘‘Insight Toolkit’’ has been developed and makes available a variety of open source image processing algorithms for computing segmentation and registration of medical data. The Visible Human Web site is one of the most popular of all NLM’s Web offerings.

NLM’s Extramural Programs for more than 20 years has supported the training of medical informaticians at universities across the nation. In the early years the program focused on training of informaticians for clinical care. Today the training programs have added opportunities for training in bioinformatics, the field of biomedical computing for the large datasets characteristic of modern research. At present, NLM provides 18 grants to biomedical informatics training at 26 universities, supporting 250 trainees. NLM also participates in the NIH Roadmap activities, almost all of which have major emphasis on biomedical computing. For example, training is an important requirement of the National Centers for Biomedical Computing, an initiative for which NLM is one of the key leaders. Training as embedded in Roadmap activities is expected to become a significant complement to NLM’s traditional support of informatics training.

THE FUTURE

In its role as the world’s largest medical library, the NLM will continue to provide free access to the enormous literature of the health sciences, including even priceless historical treasures dating to the 11th century. As to the 21st century, the Library is making major contributions to the NIH Roadmap and is also applying its unparalleled collections and talents to ‘‘BIOSHIELD,’’ the Department of Health and Human Services’ effort to combat bioterrorism. The ability to apply medical knowledge to make our citizens healthy and safe is to repay the investment of the nation in medical research. In this, the National Library of Medicine can be of great help.

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Biographical Sketch

NAME: Donald A. B. Lindberg, M.D.
POSITION : Director, National Library of Medicine, NIH
BIRTHPLACE: Brooklyn, NY
DATE : September 21, 1933
EDUCATION: A.B., magna cum laude, Amherst College, 1954;
M.D., College of Physicians and Surgeons,
Columbia University, 1958;
Diplomate: Anatomic and Clinical Pathology, 1963

EXPERIENCE
1984-Present : Director, National Library of Medicine, NIH

1992-1995: Director, National Coordination Office for High Performance
Computing and Communications, Office of Science and Technology Policy, Executive Office of the President

1996-2000: U.S. National Coordinator for the G8 Global Healthcare Applications Project, Global Information Infrastructure Initiative

1969-1984: Professor of Pathology and Director, Information Science Group,
University of Missouri-Columbia

HONORS AND AWARDS : Phi Beta Kappa

: Simpson Fellow of Amherst College

: Markle Scholar in Academic Medicine

: Surgeon General's Medallion, 1989

: First AMA Nathan Davis Award for Outstanding Member of the
Executive Branch in Career Public Service, 1989

: Walter C. Alvarez Memorial Award of the American Medical Writers Association, 1989

: Presidential Senior Executive Rank Award, 1990

: Founding Fellow of the American Institute of Medical and Biological Engineering, 1992

: Outstanding Service Award of the Uniformed Services University of the Health Sciences, 1992

: Federal Computer Week=s Federal 100 Award, 1993

: Computers in Healthcare Pioneer Award, 1993

: Association of Minority Health Professions Schools Commendation, 1995

: RCI High Performance Computing Industry Recognition Award,
1995

: U.S. National Commission on Libraries and Information Science Silver Award, 1996

: Council of Biology Editors Meritorious Award, 1996

: Presidential Rank Award of Meritorious Executive in the Senior
Executive Service, 1996

: Fellow of the American Association for the Advancement of
Science, 1996

: Medical Library Association President's Award, 1997

: American College of Medical Informatics Morris F. Collen, M.D. Award of Excellence, 1997

: Johns Hopkins University School of Medicine, Ranice W. Crosby Distinguished Achievement Award, 1998

: New York Academy of Medicine Information Frontier Award, 1999

: Cosmos Club Award, 2001

: Surgeon General's Medallion, 2002

: Honorary Doctorates: Amherst College, State University of New

York, Syracuse; and University of Missouri-Columbia