

Michael Ackerman • Deena Acton • Candace Adams • Christal Adams • Marlene Adams • Richa Agarwala • Maria Aguirre Oluwatoyin Akintoye • Barbara Albright • Brian Aleibar • Eugene Aleibar • Cassandra Allen • Michael Almeida • Andre Alston Anne Altemus • Stephen Altschul • Jason Ambrogi • Susan Anderson • Dawn Arculus • Duane Arenales • Natalie Arluk • Stacey Arnesen • Richard Arnold • Alan Aronson • Jules Aronson • Suzanne Aubuchon • Vivian Auld • Ione Auston • Katarina Avnet Yvonne Ayoub • Simon Baatz • Ibironke Babalola • Joyce Backus • Evelyn Bain • Esther Baldinger • Rica Banfield • Richard Banvard • Ellen Barber • Redmond Barnes • Virgil Barrow • Ora Battle • Anne Baxter • Roxanne Beatty • John Becker • Frances Beckwith • Marie Belony • Juliet Ben Ezra • Beryl Benjers • Deborah Bennett • Dennis Benson • Nadine Benton • Bernard Brenda • Donetta Best • Richard Better • Tariq Bhatti • Fabienne Bichotte • Anne Black • Dennis Black • Nancy Bladen • Diane Boehr • Mark Boguski • Angela Bolder • Cheryl Bolder • Cynthia Bolton • Ann Bornstein • Patricia Bosma • Jane Bouton

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES NATIONAL INSTITUTES OF HEALTH NATIONAL LIBRARY OF MEDICINE PROGRAMS & SERVICES 2000

Frederick Bowell • William Boyd • Regina Brailsford • Jana Brightwell • Carole Brown • Patrick Brown • Sharon Brown • Allen Browne • Stephen Bryant • Michael Bumbray • Sally Burke • Burns Susan • Burton Maureen • Butler John • Susan Buyer David Byers • Chika Byrd • Margaret Byrnes • Rebecca Cagle • Marjorie Cahn • Audrey Campbell • Kathi Canese • Phuc-Lanh Cao • Jane Carlton • Annmarie Carr • Victor Carr • Doris Carroll • Patricia Carson • Vivian Carter • Karen Carter-McLeod Dana Cartwright • Karen Casey • Barbara Cassatt • James Cassedy • Mark Cavanaugh • Dana Chan • Quinn Chan • Reginald Chandler • Karen Chappelle • James Charuhas • Joseph Charuhas • May Cheh • Lucie Chen • Meei-Ming Chen • Price Cherre Chong Chung • Deanna Church • Victor Cid • David Cissel • Karen Clark • Tamar Clarke • Carol Clausen • Peter Clepper Victor Clunet • Martha Cohn • Ann Coleman • Florine Coleman • Nelda Coligan • Maria Collins • Donald Comeau • Yhorda Conkle • Mary Conway • Mary Cooney • Kathy Cooper • Peter Cooper • Milton Corn • Nancy Cox • Nicholas Cozzarelli Kathleen Cravedi • Robert Cross • Sharon Cummings • Walter Cybulski • Twann Dailey • Isamu Danura • Noveen Das Laurence Dash • Frank Davidoff • Sara Davidson • Lynn Davis • James Dean • Peter DeAnna • Amanda DeCalderon • Stephen DeFreytas • Andrea Demsey • Felicia Derricott • Richard Desper • Michael Detweiler • Victoria Detweiler • Monika Devay Eugene Dickinson • Patricia Dickinson • Jean Diehl • Lanetta Dietrichson • Erin Dillenbeck • Phuong Do • Yoshiko Doherty Nancy Dosch • Tamas Doszkocs • Simmie Douthit • Walter Dudley • Mary-Kate Dugan • Jacqueline Duley • Erinn Dumas Laurie Duquette • Cecilia Durkin • Gale Dutcher • Judith Eannarino • David Eckstein • Erica Edding • Jacqueline Egbaiyelc Linda Eisenstadt • Andrea Epstein • Lee Erlichman • John Esler • Colin Eversley • Michael Ewing • Maria Farkas • Sheila Fauble • Herbert Federhen • Elizabeth Fee • Patricia Feldman • Margaret Feng • Tyrone Ferguson • Michael Fetchko • Cheryl Fields • Martha Fishel • Albert Fisher • Barbara Fisher • Carol Fisher • Joe Fitzgerald • Jazzmyn Flemmings • Skokitha Flemmings • James Fleshman • Judith Folkenberg • George Fonger • Paul Fontelo • Kimberlee Ford • Jane Fowler • Anne Fox Garry Fox • Uta Francke • George Franklin • Reginald Frazier • Mark Freidin • Herbert Friedman • Cynthia Gaines • Marie Gallagher • Michael Galperin • Robert Galvin • Helen Garton • Lillie Gaskins • Lewis Geer • Gary Gernhart • John Gibb Michael Gill • Yvonne Gilpin • Paul Ginsparg • Adam Glazer • Martha Glock • Ronald Gordner • Tameka Gore • Jeanne Goshorn • Charlet Grant • Stephen Greenberg • Jane Griffith • Karen Guarin • Vivien Guckenheimer • Carol Haberman • Karen Hajarian • Kathi Hanrahan • Mary Hantzes • Charles Harbert • Dwayne Hardware • Alvin Harris • Nakita Harris • Daniel Hartinger • Susan Hauser • Miranda Hay • George Hazard • Belinda Hazell • Kimberly Hazell • Earl Henderson • Jacqueline Henrich • John Henrich • Darnell Henson • Malvin Henson • Charles Herbert • Mary Herron • Daniel Higgins • Kenneth Hill Mazzateria Hill • Ruth Hill • Troy Hill • The Ho • Colette Hochstein • Christa Hoffmann • William Hole • Nicole Hollis • Crysta Holman • Michael Homan • Meredith Horan • Carol Hotton • Dorcus Howard • Myrtis Howard • Fu-Sen Hu • Min-Chi Huang Vera Hudson • Susanne Humphrey • Betsy Humphreys • Jennie Hunt • Randy Huntzinger • Joseph Hutchins • Margaret Hutto Sithipong Intorant • Aravind Iyer • Britonya Jackson • Karlton Jackson • Alice Jacobs • Maile Jacobsen • Wonhee Jang Donald Jenkins • Mary Jenkins • Cheryl Johnson • Frances Johnson • William Johnston • Barbara Jones • Ethel Jones • Avior Jordan • Judy Jordan • Tonya Judy • Lora Jurgrau • Sanita Kacker • Margaret Kaiser • Jonathan Kans • Wen-Min Kao • Judith Kaplan • Bonnie Kaps • Jane Karkashian • Deborah Katz • Kenneth Katz • David Kenton • Sunnie Kim • Pandoria King Lawrence Kingsland • Paula Kitendaugh • Paul Kitts • Lori Klein • Lou Knecht • James Knoblen • Rita Koegel • Alexey

NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE

PROGRAMS AND SERVICES
FISCAL YEAR 2000

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
BETHESDA, MARYLAND

National Library of Medicine Catalog in Publication

Z National Library of Medicine (U.S.)
675.M4 National Library of Medicine programs and services.--
U56an 1977- . -- Bethesda, Md. : The Library, [1978-
v. : ill., ports.
Report covers fiscal year.
Continues: National Library of Medicine (U.S.). Programs and services. Vols. for
1977-78 issued as DHEW publication ; no. (NIH)
78-256, etc.; for 1979-80 as NIH publication ; no. 80-256, etc.
Vols. for 1981-available from the National Technical Information Service,
Springfield, Va.
ISSN 0163-4569 = National Library of Medicine programs and services.

1. Information Services - United States - periodicals 2. Libraries, Medical -
United States - periodicals I. Title II. Series: DHEW publication ; no. 80-256, etc.

DISCRIMINATION PROHIBITED: Under provisions of applicable public laws enacted by Congress since 1964, no person in the United States shall, on the ground of race, color, national origin, sex, or handicap, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance. In addition, Executive Order 11141 prohibits discrimination on the basis of age by contractors and subcontractors in the performance of Federal contracts. Therefore, the National Library of Medicine must be operated in compliance with these laws and executive order.

Contents

Preface	v
Office of Health Information Programs Development.....	1
Long Range Plan	1
International Programs	2
Outreach Activities.....	4
Library Operations	6
Planning and Management	6
Collection Development and Management	7
Bibliographic Control.....	9
Information Products.....	12
Direct User Services.....	16
Outreach	17
Health Informatics Activities	22
Specialized Information Services	26
Resource Building	26
Resource Access.....	28
AIDS Information Services	29
Outreach/User Support.....	29
Lister Hill Center	31
Introduction	31
Knowledge Processing	32
Information Systems	36
Image Processing.....	39
Medical Informatics Training.....	44
Office of the Public Health Service Historian.....	44
Resource Support and Development	44
Engineering Laboratories	46
External Research Support.....	47
National Center for Biotechnology Information	50
GenBank: The NIH Sequence Database	50
The Human Genome	52
PubMed	55
The BLAST Suite of Programs	56
Other Specialized Databases and Tools	57
Database Access.....	59
Basic Research	62
User Support	63
Outreach and Education	64
Extramural Programs.....	65
Biotechnology Information in the Future.....	65
Extramural Programs	66
Resource Grants	66

Training and Fellowships	67
Research Support.....	68
Other Grants	69
Special Projects	70
Grants Management Highlights	71
Summary	72
Office of Computer and Communications Systems	73
Overview	73
Customer Services.....	75
Desktop Support.....	76
Network Support	77
System Support	79
System Security.....	80
Computer Facilities	81
Reinvention Systems.....	82
Administrative Support Systems.....	88
Administration.....	90
National Performance Review	90
Financial Resources.....	91
Personnel.....	91
NLM Diversity Council	101
NLM Organization Chart	(inside back cover)

Appendixes

1. Regional Medical Libraries	102
2. Board of Regents	103
3. Board of Scientific Counselors/LHC.....	104
4. Board of Scientific Counselors/NCBI	105
5. Biomedical Library Review Committee	106
6. Literature Selection Technical Review Committee.....	108
7. PubMed Central National Advisory Committee.....	109

Tables

Table 1. Growth of Collections.....	23
Table 2. Acquisition Statistics	23
Table 3. Cataloging Statistics	24
Table 4. Bibliographic Services.....	24
Table 5. Circulation Statistics.....	24
Table 6. Online Searches—All Databases.....	25
Table 7. Reference and Customer Service.....	25
Table 8. Preservation Activities.....	25
Table 9. History of Medicine Activities	25
Table 10. Extramural Grants and Contracts Program.....	72
Table 11. Financial Resources and Allocations	91
Table 12. Full-time Equivalents (Staff)	101

Preface

The Library continues its conversion from an institution aimed solely at serving the community of health professionals—practitioners, scientists, educators—to an institution devoted to bringing science-based health information to all. There have been several notable advances this year in that endeavor:

- MEDLINE*plus*, introduced in October 1998, continues to be expanded in scope. There are now more than 400 “health topics” on medical subjects of widespread interest to the general public. Two important additions to MEDLINE*plus* this year were an extensive medical encyclopedia, with thousands of illustrations, written in lay language, and, through a special arrangement with the U.S. Pharmacopeia, detailed information about more than 9,000 brand name and generic prescription and over the counter drugs. The popularity of this service may be seen in its current rate of usage: 5 million page hits a month.
- A signal accomplishment this year was developing and making available to the public a new database, ClinicalTrials.gov. Created by professional staff in the Lister Hill Center, the new database went up on NLM’s web site in February 2000 and is already logging more than 2 million page hits a month. ClinicalTrials.gov contains more than 5,000 studies in 50,000 locations. We believe it will be of great help to both physicians and patients.
- The “interface” to MEDLINE, known as PubMed, was substantially improved in 2000, including the addition of easy options to introduce limits to searches, to sort references, and to view an alphabetic list of related terms; a “history” feature that keeps track of everything you’ve done in a MEDLINE session; and a “clipboard” feature that allows you to collect, view, save, and print selected citations from one or several searches. The PubMed system also has (as of September 2000) links to more than 1100 participating publishers’ Web sites so that users can retrieve full text versions of articles identified in a MEDLINE search.
- The NLM funded 53 “outreach” projects around the Nation in FY2000. The purpose is to increase Internet access to good health care information in a variety of settings—public libraries, middle schools, malls, senior centers, etc. Many of the projects span rural, inner city, and suburban areas.

These are but a few of the advances you will discover in this report. They are the result of hard work by a superb, dedicated staff. Our thanks go also to the Regents and other advisors who serve on the Library’s committees.

Donald A.B. Lindberg, M.D.
Director

OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

Elliot R. Siegel, Ph.D.
Associate Director

Long Range Plan

The NLM Board of Regents has published an *NLM Long Range Plan 2000-2005* for the new century. NLM's original 1987 Plan and subsequent supplements have served as a strong basis for planning, management, and resource allocation. At its January 2000 meeting, the Board approved a new NLM Long Range Plan for 2000-20, developed with the advice and participation of over one hundred past planning panel members and other consultants. The new Plan grows out of the Library's very successful fifteen-year history of long range planning.

The Plan identifies four overall goals for NLM:

- Organize health-related information and provide access to it
- Encourage use of high quality information by health professionals and the public
- Strengthen the informatics infrastructure of biomedicine and health
- Conduct and support informatics research

Within the structure of these four goals, the Plan identifies seven recommended priorities for special emphasis in addition to support of basic library services:

1. Health information for the public

NLM has historically focused its services and products on an audience of health professionals and biomedical scientists. With widespread deployment of computers and telecommunications, the time is now right for NLM to provide access to health information that is useful both to the general public and to practitioners who need information outside their particular field of expertise. The Plan also recommends that NLM promote research on ways that information services can improve personal health care decisions and outcomes.

2. Molecular biology information systems

The Plan recommends that NLM continue its commitment to organize genomic data to meet the rapidly evolving genome research agenda. The explosive growth in the fields of genetics and molecular biology, spurred largely by the worldwide success of the Human Genome Project, has resulted in staggering volumes of data that have increased by many orders of magnitude over the past decade. The challenge for the next decade will be to keep pace with the flood of genome data, while also designing the tools and databases for the gene discoveries of the 21st century—discoveries that will advance understanding of molecular processes affecting human health and disease.

3. Training for computational biology

The nation's biomedical research enterprise needs more trained professionals in computational biology, including mathematical modeling in the life sciences, imaging and molecular biology. NLM should contribute to NIH efforts to increase the number of people who are trained in computational biology, by building on its unique informatics training program that bridges the gap between basic and clinical research.

4. Definition of the research publication of the future

NLM should play an active role in defining the research publication of the future. Electronic methods for disseminating biomedical research results (such as PubMed Central) are being developed, keeping pace with the rapid improvements to electronic computing and communications technologies. As a major player in the management of scientific information, NLM should contribute to the development of new forms of publishing which can provide more rapid exchange of information, increased multimedia capabilities, the opportunity for lower dissemination costs, and wider global accessibility.

5. Permanent access to electronic information

The rapid increases in electronic publishing and technological change make the problem of ensuring long-term access to electronic

information difficult NLM must be a leader in responding to the problem of impermanence of electronic information. As a creator, organizer, and disseminator of information in electronic form, NLM has a responsibility to contribute to the development of technical methods and affordable collaborative strategies. Success will require collaboration with other libraries and a range of stakeholders to develop the necessary technical standards, and scalable national and international approaches required to ensure permanent access.

6. Fundamental informatics research

The Plan notes that advances in computing, storage, and communications provide new opportunities for productive basic research in medical informatics and digital libraries. NLM should increase resources for extramural and intramural research in these areas. A major problem for research is how to build robust systems that tailor “just in time” answers to specific questions that occur to busy clinicians in the context of direct patient care. The Plan suggests that NLM should explore the potential of research and development in information systems that move beyond information retrieval to provide specific knowledge needed for clinical decision-making. A related research issue is how to help patients and families find information specific to their immediate health concerns.

7. Global health partnerships

The increasing globalization of knowledge has made it clear that domestic and international functions of the NLM are not separable. The Plan reaffirms the international mission of NLM. In particular, the Library should focus on establishing new partnerships to leverage its resources. It should also seek to improve the effectiveness of the international initiatives of others (e.g., health science centers and libraries, research funders, donor organizations, non-governmental organizations, etc.) through improved access to and use of new computer and information technology and knowledge management tools. It is important that NLM carefully select targets of opportunity for involvement in areas of the world where NLM can make a difference.

International Programs

Internet Connectivity at Malaria Research Sites in Africa

NLM continues to lead the Communications Working Group of the Multilateral Initiative on Malaria (MIM), which was begun in 1997. The objective of MIM is to allow African scientists and malaria researchers to connect with one another and sources of information through full access to the Internet and the resources of the World Wide Web, as well as the creation of new collaborations and partnerships.

The initial meeting of the MIM CWG was held in January 1998 at the NLM. In attendance were malaria research scientists, health information professionals, telecommunications experts and representatives of the major MIM funding agencies. In keeping with the underlying goal of supporting a broad spectrum of basic and operational malaria research needs, the researchers requested communications and connectivity capabilities sufficient to provide, at a minimum: robust and reliable e-mail, links to other research sites, access to full text journal articles, database searching, exchange of large files and mapping data, and timely access to electronic information resources worldwide.

Accomplishments:

- Brokered an arrangement with a VSAT vendor in which the vendor gives NLM a group rate for the MIM sites and is treating the group of sites as one customer.
- The group arrangement has advantages. The rate per site would increase considerably if the site were to buy it on its own. In addition, the consortium approach allows for flexibility in adjusting bandwidth to fit the needs of the individual sites.
- NLM supports site visits and assessments, design of implementation strategy for each site, advice on in-country licensure of technology, oversight of installation as well as continued support and training.
- NLM recommends immediate use be made of the affordable technologies now available

to provide high-speed and reliable information and communication links in order to yield timely results in improving researchers' ability to do co-operative research and disseminate their results. Recommended technologies are VSAT, which uses a geostationary satellite and an earth station, and microwave, which uses radio waves. The latter is less expensive but is limited to line of sight transmission.

- In July 1999, NLM successfully installed VSAT ground stations at two malaria research sites in Kenya at Kisian (CDC funded) and Kilifi (Wellcome Trust funded). These two sites join the Malaria Research and
- Training Center in Mali which has full Internet access via microwave technology, funded by NIAID and made operational in June 1998.
- In December 1999, the NLM team brought on two additional sites in Ghana—in Accra (Noguchi Institute) and in Navrongo (Navrongo Health Research Center). The Ghanaian sites, engaged in malaria vaccine testing, will be funded jointly by NIAID/NIH, the Naval Medical Research Center, and USAID.) The overall bandwidth for the network was increased to 128kbs and monthly charges are \$2,100/month per site.
- In March 2000, NLM installed a VSAT station in Nairobi at WRAIR headquarters. This station also serves CDC, Wellcome Trust, the Kenyan Medical Research Institute (KEMRI), and the Library of Congress.
- In September, 2000, NLM installed two stations and local area networks at remote research sites in Tanzania as part of a collaboration with the Tanzanian National Institute of Medical Research (NIMR): One at Ifakara Center and another at Amani. At NIMR headquarters in Dar es Salaam, NLM installed a microwave link to the local ISP. These installations were funded by OD/NIH.
- In October 2000, NLM sponsored a comprehensive training workshop for all IT personnel from the malaria research sites in Africa.

Regulatory authorities within each country have control over the installation and use of all communications technology. In some instances, this does not present a problem; in other instances, Kenya, for example, it is a stumbling block.

International DOCLINE Libraries

Under an NLM pilot project that began in 2000, malaria researchers in Southern, Central and East Africa have been able to request malaria-related documents and journal articles they need through the medical libraries at the University of Zimbabwe and the South African Medical Research Council. Their requests are filled either locally by one of these African library sites or at the NLM as quickly as possible—by Internet, e-mail, fax, or airmail. This service is provided by NLM free of charge and is intended to serve malaria researchers with a range of technological capabilities:

- World Wide Web application: Under this service, malaria researchers with web access can search MEDLINE via the NLM's free PubMed service and request the articles electronically using the Loansome Doc feature. Note that some journals are linked full-text to PubMed references and can be obtained directly online (e.g., *British Medical Journal*).
- E-mail application: For malaria researchers who would like to use e-mail as a document delivery method but who do not have document image application software (viewer), a free software application called DocView, developed by the Communications Engineering Branch of NLM, is available. DocView enables the user to view, zoom, shrink, scroll, pan, rotate, and print bitmapped image documents received via email.
- Fax and Air Mail application: Documents can also be faxed or air mailed to researchers as the preferred method of delivery, especially for researchers without reliable e-mail connections or who prefer paper copies of documents.

Special Malaria Web Site: Malaria researchers with WWW access are invited to access a web site of useful resources and links for malaria research at www.mimcom.net

Global Internet Connectivity

In 1999–2000, NLM completed its first phases of end-to-end connectivity testing and evaluation. This included further exploration of the methods and metrics needed to better understand the quality of Internet performance from the end-user perspective. NLM collaborated with numerous domestic and international partners to test Internet pathways in the U.S. and around the world. Pathways included links between the U.S. and North, Central, and South America; Western and Eastern Europe; the Middle East; sub-Saharan Africa; and South, Southeast, and East Asia.

During 2000, NLM extended the Internet testing to include high bandwidth Internet connections, including the vBNS (very high Bandwidth Network Services). NLM conducted pilot tests in the U.S. as a prelude to commencing tests of international high-bandwidth connectivity. NLM is particularly interested in determining if the pilot test results apply globally. Briefly, the pilot tests concluded that simply switching to a high-bandwidth connection did not in and of itself result in significantly improved performance. In order to realize the benefits of high-bandwidth connections, the operating system parameters needed to be changed or “tuned” to optimize use of the greater bandwidth. With properly tuned systems, performance does improve markedly.

Outreach Activities

The NLM has a longstanding commitment to the effective dissemination and use of biomedical information within the health community. To help achieve this goal, NLM has, since 1989, collaborated with its National Network of Libraries of Medicine (NN/LM) to conduct outreach to health professionals and especially those in rural, minority, or other underserved communities. The objectives of NLM-sponsored outreach are to: 1) make health professionals aware of the information products

and services NLM provides; 2) facilitate health professionals’ access to and use of biomedical information; 3) provide training in the searching of electronic databases; 4) assist health professionals in developing new information-seeking behaviors; and 5) improve health care practices through the use of authoritative, up-to-date information.

Between 1990 and 1995, NLM supported close to 300 outreach projects that involved more than 500 institutions across the country. In 1996, NLM published a 5-year review of its outreach program and activities. The review concluded that NLM’s outreach program has made significant progress overall. However, the review also recommended that the methodology for evaluating outreach be more fully developed. The 5-year review envisioned that strengthened outreach evaluation would help NLM and the NN/LM more clearly discern lessons learned from past experience, better plan for future outreach activities, and design future outreach with a built-in evaluation component to the extent feasible.

NLM determined that a logical next step in outreach would be to undertake a special project to develop a framework or model of outreach planning and evaluation. The working hypothesis is that the medical library community would benefit from knowledge of evaluation studies of outreach-like activities that have been carried out in related disciplines. NLM is especially interested in exploring related fields which have evaluated efforts directed toward minority populations because outreach to minority and other underserved populations is one of NLM’s highest priorities, and, at the same time, an area in which success has been most difficult to achieve.

The evaluation outreach project has now attained most of its original objectives. The project was a collaboration between NLM’s Office of Health Information Programs Development and NN/LM’s Pacific Northwest Regional Medical Library at the University of Washington. The project used an interdisciplinary advisory panel to provide input and advice in the drafting of an outreach planning and evaluation field manual. The field manual went through several drafts and panel reviews, followed by outside review and field

testing by faculty, students, and practitioners in the library and information sciences.

The field manual was published in September 2000, and is available on the web at www.nlm.nih.gov/evaluation. NLM encourages broad use of the field manual by health information outreach planners, practitioners, evaluators, researchers, and teachers. Staff at the Pacific Northwest RML and at NLM are available to advise on follow-up activities that involve application of the field manual to relevant activities. The field manual is already being used in a variety of NLM-sponsored projects.

A separate but related project is providing improved Internet connectivity and related technical support and training to Native Americans in the Pacific Northwest and Pacific Southwest. At present, NLM has supported tribal connections projects with 16 American Indian tribes and Alaska Native villages in the northwest, and 4 American Indian tribes in the southwest. Phase 1 (Pacific Northwest) of tribal

connections is substantially complete, with final project evaluation now under way. Phase 2 (Pacific Southwest) sites have been selected, and implementation is beginning.

NLM/OHIPD is supporting a special tribal connections project in the Southeast, with the American Indian Cultural Center/Piscataway Indian Museum in Waldorf, MD. Here, NLM has supported the installation of a computer lab and computer resource learning center, and relating training and evaluation. The ribbon-cutting ceremony for the computer infrastructure was held on November 15, 2000, with the participation of Dr. Yvonne Maddox, NIH Acting Deputy Director and Dr. Donald A.B. Lindberg, NLM Director. The event was covered by Maryland Public TV. Combined with the other tribal connections projects, NLM is developing a robust knowledge base for outreach to Native American and other underserved communities, especially those in rural and remote areas.

LIBRARY OPERATIONS

Betsy L. Humphreys
Associate Director

NLM's Library Operations (LO) Division provides the basic services that ensure access to the scholarly record of biomedicine and the health professions. LO selects, acquires, preserves, and organizes the world's biomedical literature; maintains a thesaurus and a classification used to organize biomedical information; produces authoritative indexing and cataloging records; builds and disseminates bibliographic, directory, and full-text databases; provides national back-up document delivery, reference, and research assistance; helps health professionals, researchers, librarians, and the general public to make effective use of NLM's services; and coordinates the 4,500 member National Network of Libraries of Medicine (NN/LM) which delivers health sciences library services throughout the country. The basic information services provided and coordinated by LO are the essential foundation for NLM's outreach programs to health professionals and the general public, as well as for its biotechnology, AIDS, and health services research information programs.

The largest of NLM's Divisions, LO employs a multidisciplinary staff of librarians, technical information specialists, subject experts, health professionals, historians, and technical and administrative support personnel. In addition to providing basic library services, the LO staff directs the National Information Center on Health Services Research and Health Care Technology (NICHSR); carries out an active program in the history of medicine; works with other NLM program areas to develop new and enhanced products and services; conducts research and evaluation studies related to the Library's programs and services; directs and participates in research in advanced information storage and retrieval; directs a post-graduate training program for medical librarians; and contributes to the development of standards for health data and knowledge-based information. LO staff members participate actively in Library-wide efforts to improve the quality of

work life at NLM, including the Diversity Council and the NLM Intranet.

Planning and Management

In FY 2000, LO planning efforts focused on assisting the development of two Library-wide plans: the *NLM Long Range Plan, 2000-2005* and the closely related *NLM Strategic Plan to Reduce Racial and Ethnic Health Disparities, 2000-2005*. NLM's overarching goals and objectives are directly reflected in LO's program priorities, which include improving access to health information for the general public and developing effective strategies for organizing, preserving, and providing access to "born digital" information. Information about LO's activities related to health information for the public and electronic publications appears throughout this report.

During the past year, LO devoted considerable management attention to three key elements of the infrastructure for basic services:

- the automated systems used in internal operations and user services;
- the space needed for the NLM collection, onsite users, and staff; and
- the contracts that support the National Network of Libraries of Medicine.

LO continued to work closely with the Office of Computer and Communications Systems (OCCS) and other NLM program areas on the multi-year project to replace essentially all of NLM's legacy systems and to end the Library's reliance on mainframe computers. In FY2000, NLM completed the implementation of the new system for creation and maintenance of the Medical Subject Headings (MeSH); installed a new web-based DOCLINE system (including new versions of SERHOLD and DOCUSER) to support automated document request generation and routing in the NN/LM; and implemented the first phase of a new indexing Data Creation and Maintenance System (DCMS). Substantial progress was made on the complex task of identifying and moving unique journal citations and unique records for monographs and chapters from specialized databases, e.g., AIDSLINE, HISTLINE, BiothicsLINE, POPLINE, into the

PubMed database and the LOCATOR^{plus} online public access catalog respectively.

In FY 2000, NLM initiated formal planning for a new building and renovation of the two existing buildings to accommodate growth of the collection and programs, including the National Center for Biotechnology Information, that were established after the Lister Hill Center building was constructed. LO provided extensive documentation of space requirements for various elements of the collection, public service areas, exhibitions, and LO staff and onsite contractor personnel. To make more effective use of existing space, LO and Lister Hill Center staff organized the move of contractor staff responsible for two of the indexing data entry streams (scanning/optical character recognition and editing of electronic data supplied by publishers) from the Lister Hill Center building to the Federal building in downtown Bethesda. Space on the B3 level was remodeled to accommodate the historical picture and film collections and the staff who manage and provide access to these collections. LO continued its practice of conducting ergonomic evaluations of workspaces and making adjustments as needed to prevent or alleviate staff discomfort.

The request for proposals for new 5-year contracts for the Regional Medical Libraries in all eight regions of the National Network of Libraries of Medicine was issued in March 2000. The statement of work for 2001–2006 emphasizes outreach to the public, as well as services to network members and outreach to health professionals, and encourages cross-regional collaboration. There is also additional attention to mid-course evaluation of RML effectiveness. In addition to the existing National Online Training Center, the solicitation invited proposals for an Outreach Evaluation Center and a Mapping Center. Proposals were received in June and the initial technical reviews were conducted in August by four review teams, each consisting of an academic health sciences librarian, a hospital librarian, a health professional, and an NLM staff member. Questions identified during these reviews have been sent to each bidder. The process of selecting the successful bidders and awarding the new NN/LM contracts will be completed in

the spring of 2001, after site visits in some Regions.

Collection Development and Management

Many NLM services depend on the Library's comprehensive collection of biomedical literature. LO ensures that NLM's collection meets the needs of current and future health professionals and researchers by developing and updating NLM's literature selection policy; acquiring and processing literature that meets this selection policy as well as electronic resources responsive to the general public's need for health information; organizing and maintaining the collection for efficient current use; and preserving materials for future generations. At the end of FY2000, the NLM collection included 2.3 million volumes and 3.6 million other items, including manuscripts, microforms, pictures, audiovisuals, and computer software.

Selection

LO staff and agents select literature for the NLM collection following guidelines in the *Collection Development Manual of the National Library of Medicine* (www.nlm.nih.gov/pubs/cdman.pdf), which typically undergoes a major review and revision every 5 to 10 years. In between major revisions, LO develops operational guidelines for materials in emerging disciplines or in formats or subjects that are posing selection difficulties. In FY2000, LO participated in the NLM-wide effort to complete a formal NLM *Policy on Acquiring Copyrighted Material in Electronic Format* (www.nlm.nih.gov/pubs/acqcopyrightmat.html). Under this policy, LO licenses electronic materials only under conditions that do not impede its ability to serve as the national backup collection in biomedicine and related fields.

In FY2000, responsibility for selection of contemporary literature was consolidated in the Selection and Acquisitions Section of the Technical Services Division (TSD). Previously the Serial Records Section was responsible for selection of new serial titles. Work proceeded on expanding the selection and acquisition of health policy "gray" literature. LO reviews segments of

the collection periodically to determine that extent to which it has been successful in adhering to selection guidelines in particular subject fields. The Library's book collections in physics and chemistry were reviewed this past year, and some items were weeded from the collection as a result.

Acquisitions

TSD received and processed 166,020 contemporary books, serial issues, audiovisuals, and computer software packages. (Table 2). Net totals of 44,208 volumes and 173,931 other items (e.g., manuscripts, microforms, pictures, audiovisuals) were added to the NLM collection in FY2000. Acquisition production levels returned to normal levels, after an expected temporary decline last year during the transition from legacy processing systems to the Voyager Integrated Library System. In FY2000, TSD expanded its use of Voyager capabilities for invoice processing, fund accounting, and electronic data interchange (EDI) for serials invoices. TSD also participated in the ILS vendor's Acquisitions Task Force to design improved functionality for future releases of Voyager's Acquisitions module.

In October 2000, TSD made new awards to vendors for the supply of foreign and domestic monographs and related services. Two new arrangements were established to supply medical Africana from countries not covered by the Library of Congress Overseas Acquisitions Program in Nairobi. NLM acquires U.S. government publications as a selective depository library under the Government Printing Office's Federal Library Deposit Program. In FY2000, TSD completed the extensive self-study report of its handling of depository materials that is periodically required of all participants in the program. An increasing number of U.S. government publications are available electronically, and NLM links directly to the electronic version from its online catalog. TSD is also expanding acquisitions of copyrighted electronic materials in accordance with the new NLM policy for this category of material. In FY2000, the emphasis was on journals available only in electronic format and resources needed for incorporation into the

MEDLINEplus consumer health information service.

The History of Medicine Division (HMD) continued to build the Library's outstanding holdings of early printed books, manuscripts, pictures, and historical audiovisuals. Notable individual works acquired in FY2000 included: Isaac Israeli's *De Particularibus Dietis Libellus* (Padua, 1487), a book on diet written in the 10th century by a Jewish physician; Giovanni Battista Carcano Leone's *Anatomici Libri II* (Ticini, 1574), a work on the anatomies of the lachrymal duct and the fetal heart; Thomas Sydenham's *Methodus Curandi Febres* (London, 1666); Baro Urbigerus' *Aphorismi, oder Gewisse Regeln...* (Erfurt, 1690?), a rare collection of alchemical aphorisms on elixirs with recipes for the distillation; and Alfred Russell Wallace's *Contributions to the theory of natural selection* (London, 1870), the first edition of his book on the principles of evolution. Manuscripts acquired include: 80 dissertations from an early New York state medical school, the College of Physicians and Surgeons of the Western District, covering the 1820s and 1830s; the first 50 years of records of the Group Health Association, which was founded in the 1930s; a large collection of oral histories related to dermatology from a project sponsored by the Dermatological Foundation of Miami; oral histories from the U.S. Food and Drug Administration; and additions to existing collections of Nobel prize winning scientists. NLM's fine collection of public health posters was enriched by the addition of many contemporary posters from around the world, as well as two nineteenth century French posters on the evils of alcoholism, a striking poster "Exposition d'Hygiene" designed by A. Graesner for a 1935 exposition in Strasbourg, and a 1940s American poster promoting polio research. Additions to the historical film collection included: 400 videotapes related to NIH and other biomedical activities from independent filmmaker Jason Vogel and videotapes of the 1986 NIH AIDS Conference and a 1986 interview with Anthony Fauci donated by Lucy Jarvis.

Preservation and Collection Management

To preserve NLM's archival collection and keep it accessible and in good order, LO carries out a range of preservation and collection management activities. These activities include binding, microfilming, conserving rare and unique items in their original formats, maintaining appropriate storage conditions and facilities for all types of material in the collection, and preventing and responding to emergencies that could damage library materials. LO distributes data about what NLM has preserved in order to avoid duplication of effort by other libraries and provides preservation information useful to other libraries on NLM's web site. NLM monitors developments in preservation technology and conducts experiments with additional preservation techniques as warranted. The Library also continues to promote the use of more permanent media in new biomedical publications.

In FY2000, LO bound 31,874 volumes, microfilmed 4,513 volumes, repaired 2,000 volumes in the onsite book repair and conservation laboratory, and conserved 385 items from the historical collection, including *Treatises on medicine*, the 12th century manuscript that was returned to the Library in FY1999 after being missing from the collection for about 50 years. The Preservation and Collection Management Section of the Public Services Division completed the implementation of Voyager Integrated Library System functionality for microfilming and collection management functions (with the exception of binding, which will not be available until a future release of Voyager). The backlog of Voyager record creation and update for preservation microfilms and recently bound journal volumes was eliminated, as was the backlog of microfilm masters to be prepared and shipped to offsite storage.

Collection management projects completed in FY2000 included: the shift of 1985–89 serial volumes from the B1 to the B3 levels; the shift of some parts of the historical collections to accommodate the consolidation of the image collections and related staff members on B3; and the inventory of 1800–1914 printed

works. The multi-year project to repair or replace aging compact shelving continued. In September 2000, a random sample of the post-1800 collection was conducted to determine how many volumes would benefit from deacidification, i.e., how many are acidic but not yet brittle. Preliminary results show that an estimated 65,832 additional volumes have become brittle since the 1985 survey; an estimated 516,600 volumes could be treated by current deacidification methods and an additional 261,300 volumes on acidic, slightly glossy paper may benefit from new deacidification processes now under development. NLM will conduct a test of the Bookkeeper® deacidification system in FY 2001.

During FY2000, an NLM-wide working group chaired by the Preservation and Collection Access Section defined and tested a set of permanence ratings to be used to designate the Library's commitment to providing permanent access to its own electronic publications. The group first identified three core categories of permanence (identifier validity, resource availability, and content invariance) and then refined through testing four permanence ratings that seem to encompass most of NLM's electronic output. These are: "Permanent: Unchanging Content" (e.g., an image of original correspondence in *Profiles in Science*); "Permanent: Stable Content" (e.g., a MEDLINE record, which may undergo changes and updates, but only of a predictable and limited kind); "Permanent: Dynamic Content" (e.g., the NLM Home Page); and "Permanence Not Guaranteed" (e.g., preliminary agenda). In FY2001, NLM will develop an operational test of the technical and administrative procedures needed to implement permanence ratings for the Library's electronic publications and to ensure that the stated permanence commitments are met.

Bibliographic Control

To facilitate access to the biomedical literature, LO creates authoritative indexing and cataloging records for journal articles, books, films, pictures, manuscripts, and electronic media. Like many others in the library and

information science community, LO is experimenting with modifications to its standard indexing and cataloging record formats for organizing and enhancing access to electronic resources. LO also maintains the Medical Subject Headings (MeSH) used by NLM and many other institutions to describe the subject content of biomedical information; collaborates with the Lister Hill Center to produce the Unified Medical Language System® (UMLS®) Metathesaurus®, of which MeSH is an important component; and maintains the *National Library of Medicine Classification*, a scheme for arranging physical library collections by subject that is used by health sciences libraries around the world.

Thesaurus Development

The 2001 edition of MeSH contains 19,942 main headings, 82 subheadings or qualifiers, 128 publication types, and more than 115,00 supplementary records for chemicals and other substances. Changes made for the 2001 edition include 184 new descriptors, updated names for 42 main headings, and 222 new cross-references. The Cells tree was reorganized to group the various components of a cell under the term Cell Structures and 27 new cell structure terms were added. Special efforts were also undertaken to revise and enhance the vocabulary related to pharmacologic actions, viruses, vaccines, and alternatives to animal testing.

The examination of the MeSH vocabulary related to alternatives to animal testing was part of a larger project which NLM undertook on behalf of the National Institutes of Health. The MeSH Section convened an ad hoc advisory of external experts and NIH personnel to examine alternatives for improving retrieval of articles about alternatives to animal testing. In addition to enhancements to MeSH, the advisory group recommended additional journals to be indexed in MEDLINE and experimentation with the development of search filters.

The majority of the content editing for the 2001 version of the UMLS Metathesaurus was completed in FY2000 under MeSH Section supervision. The 2001 version of the Metathesaurus is expected to have about 60,000

more concepts and 400,000 more concept names than the 2000 edition.

Cataloging

LO catalogs the biomedical literature acquired by NLM both to document what is available in the Library's collection and to provide cataloging records that can be used by other health sciences libraries to reduce the level of effort required to organize their own collections. LO also catalogs or otherwise organizes information resources published on the World Wide Web, both to expand existing services, such as MEDLINE*plus*, and to conduct experiments to contribute to the development of a national approach to organizing credible web-based health information.

In FY2000, TSD's Cataloging Section cataloged 20,067 contemporary books, serials, nonprint items, and cataloging-in-publication galley, using a combination of inhouse staff and contractors. This was a 39% increase from the previous year, when production was negatively affected by the transition from NLM's legacy cataloging system to the Voyager Cataloging Module. On September 15, NLM catalogers produced their first Electronic-Cataloging-in-Publication (E-CIP) records. E-CIP takes advantage of the web environment to transmit galley and pre-publication cataloging records. The electronic routing and tracking capabilities developed by the Library of Congress (LC) reduce the time information is in transit, thus making cataloging data available more rapidly. NLM has collaborated with LC to produce CIP records for biomedical materials since the 1970s.

The Cataloging Section chaired an NLM-wide working group charged with developing a standard minimum set of metadata for the Library's own electronic publications, with reference to existing NLM metadata specifications, such as those used for *Profiles in Science*, and to NLM efforts to develop the permanence ratings described in a previous section of this report. The draft NLM metadata set is based on the Dublin Core, but is more prescriptive. (NLM's position is that, in its current format, the Dublin Core does not do enough to promote the creation of comparable metadata by different organizations.) The

Library will carry out additional testing of its draft minimum metadata set in FY2001, in conjunction with the operational test of technical and administrative procedures for assigning permanence ratings to NLM's electronic publications and ensuring its permanence commitments are met.

HMD cataloged 49 historical monographs, 256 pictures, and 49.8 linear feet of manuscripts, achieving a substantial increase in picture and manuscript cataloging. The Digital Manuscripts Program staff made substantial progress in expanding the content of Profiles in Science, (<http://profiles.nlm.nih.gov/>) a joint project of the Lister Hill Center and HMD. New sites were added for NIH Nobelists Martin Rodbell and Julius Axelrod, and the Joshua Lederberg and Martin Rodbell sites were updated. Work was substantially completed on the Christian Anfinsen site, which will be released in early FY2001.

The Cataloging Section completed work on a revision of the 5th edition of the *National Library of Medicine Classification*, which was published by the Government Printing Office in September 2000. The revised edition incorporates changes made to the schedules from 1995 to 1999 and incorporates in its index several hundred concepts added to MeSH from 1994 to 1999. As part of NLM System Reinvention, the classification database was converted to the Oracle® database management software and a web-based beta version was made available to NLM catalogers. The Library expects to make a public web version of the NLM Classification available in 2001.

Indexing

LO indexes articles from about 4,400 biomedical journals so that users of the MEDLINE database and the products generated from it can locate articles on specific biomedical topics. A combination of inhouse staff, contractors, and cooperating U.S. and international organizations perform the indexing, under the supervision of the Index Section in the Bibliographic Services Division (BSD). In addition to indexing newly published articles, LO also annotates existing MEDLINE records when the articles to which they refer have been

retracted, corrected, or challenged in subsequently published notices or commentaries.

The Literature Selection Technical Review Committee (LSTRC) (Appendix 6), an NIH-chartered committee of outside experts, advises NLM about which journals should be indexed for MEDLINE and *Index Medicus*. In FY2000, the Committee reviewed 412 journals and rated 71 sufficiently highly for immediate inclusion in MEDLINE; another 42 titles were accepted provisionally pending receipt of electronic citation and abstract data from their publishers. Several pharmacology societies reviewed MEDLINE coverage of pharmacology titles; after reviewing their recommendations, the LSTRC added 3 titles to MEDLINE. A number of journals that focus on alternatives to animal testing were added to MEDLINE based on advice from the ad hoc committee mentioned earlier. BSD worked with staff in NLM's Specialized Information Services (SIS) to identify additional toxicology journals suitable for indexing in MEDLINE as part of the SIS project to create a "virtual TOXLINE" database that eliminates duplication between MEDLINE and TOXLINE. NLM also began a cooperative project with the National Institute of Occupational Safety and Health to review and improve MEDLINE coverage of occupational safety and health journals.

In FY2000, NLM added 442,168 citations to MEDLINE, about 2% more than in FY1999. Due to increased volume caused by the higher rate of selection of journals for indexing, the substantial reduction in the number of journals that are selectively indexed, the Index Section's assumption of responsibility for indexing journals previously indexed by the American Hospital Association, and disruptions caused by implementation of the new online indexing system, a small indexing backlog developed during FY2000. BSD is taking steps to increase contractor and inhouse indexing in FY2001.

Of the citations added to MEDLINE this year, 38% were received electronically from publishers, 30% were entered via scanning and optical character recognition, and only 32% were double-keyboarded. The number of citations received electronically, the fastest and most economical method, increased 72% from

FY1999, and the number of citations keyboarded decreased 40%. At the close of FY2000, NLM was receiving XML-tagged electronic data from 181 publishers for 1,132 journals.

Regardless of the initial data entry method, all MEDLINE citations are transferred to an online indexing system where indexers add subject headings and other data elements needed to complete the citations. In FY2000, LO and OCCS began implementation of a new indexing Data Creation and Maintenance System (DCMS). Because NLM has unique indexing requirements, the new web client-server system, like its mainframe predecessor, is the result of an inhouse, custom development project. It makes use of serials data obtained from the Voyager Integrated Library System, interfaces with the MeSH browser, and receives data from the data entry streams and exports data to the PubMed database in XML format. Unlike its predecessor, the new DCMS will have a complete maintenance database of MEDLINE citations, separate from the retrieval version of the citations available to the public via PubMed. Inhouse indexers began use of the DCMS late in the fiscal year. Use will gradually be extended to NLM indexers working at home, contract indexers, and international MEDLARS centers during 2001.

LO continued to collaborate with the indexing initiative research projects, with a goal of identifying automated methods that improve access to scientific literature without increasing the level of human effort required. The new DCMS will provide a more flexible environment for giving production indexers access to automated tools that may improve the quality and quantity of their indexing, and plans for such experiments are under development.

Information Products

NLM produces online databases, other electronic resources, and print publications that incorporate its authoritative indexing, cataloging, and thesaurus data. LO collaborates with other NLM components to produce some of the world's most heavily used medical information resources.

Databases and Web Information Resources

Users conducted about 244 million searches of MEDLINE via PubMed in FY2000, with some of the searches directed to PubMed from the NLM Gateway, Internet Grateful Med, and MEDLINE*plus*. Approximately 120,000 unique "IP" addresses access PubMed each day. BSD's MEDLARS Management Section assisted the National Center for Biotechnology Information (NCBI) in designing, developing, and testing major upgrades to the PubMed system to improve searching capabilities and to implement the new LinkOut and Cubby features, which support customized links to those sources of full-text journals which have been licensed by the searcher's institution. BSD also worked with NCBI to develop and implement a strategy for allowing PubMed users to restrict their searches to various subject areas via the use of subsets and filters. A new subset was implemented for toxicology. NLM is collaborating with NIH's National Center on Complementary and Alternative Medicine to establish a PubMed subset that will improve on the functionality of the separate complementary and alternative medicine database that is currently maintained by the Center.

BSD also assisted the Lister Hill Center with the design, development, and testing of the NLM Gateway, a new search interface designed to assist users in determining which of many NLM databases contain information on particular topics of interest. The NLM Gateway does simultaneous searches of MEDLINE on PubMed, the NLM catalog in LOCATOR*plus*, consumer health information in MEDLINE*plus*, and the HSRProj health services research-in-progress database, among other resources. In addition to testing the Gateway to determine whether it was functioning as designed, the MEDLARS Management Section conducted and reported on usability testing involving intended users of the system. A number of improvements were made to the NLM Gateway based on the results of the usability testing. When remaining System Reinvention database transition tasks have been completed, the NLM Gateway will replace functionality now provided by Internet Grateful Med, and Internet Grateful Med will be retired.

Work continued on the major System Reinvention project to transfer all unique citations from NLM's specialized subject databases either to LOCATORplus, PubMed, or a new meetings abstracts database. When this process is completed, NLM will eliminate these separate databases and provide similar search functionality through the use of subject subsets and the NLM Gateway. The Cataloging Section has been working to identify, convert, and transfer unique book and chapter citations from the specialized databases and from 1970s MEDLINE data to the catalog database in LOCATORplus. The process requires substantial programming support from OCCS, as well as work in other LO Divisions. Book and chapter citations from HealthSTAR, HISTLINE, SPACELINE, and BIOETHICS are now available in LOCATORplus. Comparable citations from POPLINE and MEDLINE will be transferred to LOCATORplus in FY2001.

BSD is coordinating the effort to plan and schedule the transfer of the unique journal citations from the specialized databases into the PubMed database, which also involves substantial efforts by OCCS, NCBI, and other LO divisions. In the reinvented system, all journal citations present in PubMed must have a corresponding journal title record in LOCATORplus to serve as the source of journal authority data. As a result, the Cataloging and Serial Records Sections have been heavily involved in identifying unique journal titles not yet present in the catalog database and in obtaining and entering bibliographic data for them. The transfer of most of the unique journal citations will take place in 2001, when the new indexing DCMS is ready to support any maintenance that these citations may require.

LO continues to make progress in converting its retrospective indexing and cataloging data to electronic form. BSD added some 220,582 citations from the 1958–1959 *Current List of Medical Literature* to the OLDMEDLINE file, which now contains nearly one million records. Lakota Technologies, Inc., a Native American organization, is currently keying 1957 data. HMD made significant progress on the project to digitize the *Index Catalogue of the Library of the Surgeon General*; some of the data should become

publicly available in 2001. Plans are under way to convert all of NLM's manuscript collection finding guides to electronic archive description (EAD) format and make them available on the web.

LO assisted the Lister Hill Center in developing ClinicalTrials.gov, the new NIH clinical trials database, in several ways. BSD participated on the development team, working with NIH institutes to define appropriate data entry streams as well as design and testing of search functionality. PSD assisted with the establishment of effective links between MEDLINEplus and ClinicalTrials.gov, in time for its highly successful launch in February 2000.

In FY2000, use of MEDLINEplus, NLM's web-based consumer health information resource, increased about five-fold to 20.5 million page hits. The Public Services Division collaborates with OCCS to develop and expand MEDLINEplus, with content development and editing assistance from medical librarians at Indiana University and the University of Cincinnati. During the fiscal year, the number of health topics covered by MEDLINEplus increased from 225 to 415; links were established to and from the new ClinicalTrials.gov database (as previously mentioned); a medical encyclopedia and a drug information resource were added; and the site was redesigned based on the results of usability testing and user feedback. MEDLINEplus is routinely cited by the media and by librarians of all types as one of the most comprehensive and credible sources of health information on the web.

In FY2000, PSD worked with the NLM Office of Communications and Public Liaison to arrange for increased participation by NIH Information Officers in reviewing and providing content for MEDLINEplus topics. An NIH Advisory Group on MEDLINEplus was established which includes representatives from 8 Institutes, two members from the NIH Director's Office, the NIH webmaster, and the Deputy Director of the NIH Office of Communications and Public Liaison. A special joint project was initiated with the National Institute on Aging (NIA) to develop a related web site for seniors, based on research on

learning habits and use of technology by this population. This "Age Pages" site will be used as a training and testing environment by NLM and NIA. LO is also working with the SPRY Foundation to develop a web-based training program to assist seniors in finding health information, using MEDLINEplus as a focus.

In addition to providing a highly regarded public service, MEDLINEplus also serves as a test-bed for experiments in sharing descriptions of web resources among members of the National Network of Libraries of Medicine (NN/LM). On August 17, 2000, representatives from the University of North Carolina, Chapel Hill, the New York Academy of Medicine, and HealthWeb consortium met with PSD staff to discuss how local and regional sites can integrate MEDLINEplus links into their resources and how MEDLINEplus can link to local and regional consumer health resources. A technical working group was formed to develop record standards and a model for sharing records with these pilot sites.

The NN/LM Office contracted with the Region 6 RML at the University of Washington to develop a web-based pathfinder "Healthinfoquest" to assist public librarians in finding appropriate web resources, including MEDLINEplus, to respond to typical types of health information requests. User feedback has been very positive, and this new tool has been reviewed very favorably in the library literature.

NICHSR worked with the Lister Hill Center on changes and improvements to HSTAT (Health Services-Technology Assessment Text). NIH Clinical Center clinical trials were removed from HSTAT, where they were something of an anomaly, when ClinicalTrials.gov was released. Many new full-text documents were added to HSTAT, including evidence reports supported by the Agency for Healthcare Research and Quality, and new chapters from *Guide to Community Preventive Services*, being developed by the Centers for Disease Control and Prevention. The Lister Hill Center is developing a new interface for HSTAT which will incorporate synonym capability based on MeSH and UMLS data. The new interface will be released in FY2001 following usability testing conducted by BSD and NICHSR.

The MeSH Section collaborated with OCCS to make Stanley Jablonski's unique database of *Multiple Congenital Abnormalities Associated with Mental Retardation* available via the web, with hyperlinks to OMIM (*Online Mendelian Inheritance in Man*), to MEDLINE references, and to the MeSH Browser. Mr. Jablonski, a former Head of the Index Section, generously granted NLM permission to provide free public access to this valuable and heavily used resource.

Machine-Readable Data

NLM leases its data in machine-readable form and makes a number of its databases available via application programming interfaces (APIs) in order to promote the broadest possible use of its bibliographic and thesaurus data. Commercial companies, international MEDLARS centers, universities, and other interested organizations then make NLM data available online or in CD-ROM products, use them to improve the functionality of a variety of medical information systems, or conduct research using the data. Effective in January 2000, the Library distributes all of its data free of charge under the terms of license agreements or memoranda of understanding. MMS handles NLM's interactions with licensees of journal citations, catalog records, TOXNET data, and UMLS data.

In FY2000, 48 organizations licensed MEDLINE and other bibliographic databases. As part of System Reinvention, LO, OCCS, and NCBI completed development and testing of a new XML (extensible markup language) format for distribution of journal citations which replaces NLM's custom data distribution format. Weekly MEDLINE updates are available to licensees via file transfer protocol. The complete citation file will be distributed via DAT tape. In December 2000, NLM discontinued Z39.50 access to MEDLINE because it interfaced with the ELHILL retrieval system which was being phased out as part of System Reinvention. As there had been little use of NLM's Z39.50 MEDLINE server, the Library did not implement Z39.50 access to the PubMed system. The Entrez utilities provide an API to MEDLINE data on PubMed.

Records from the LOCATOR^{plus} catalog file are currently distributed in MARC format and are available via Voyager's Z39.50 interface. LO intends to develop an XML distribution format for its catalog records. Hundreds of organizations and individuals download MeSH data via file transfer protocol from NLM's public web site under the terms of a memorandum of understanding. MeSH data are currently available in relational files and in the MARC format. An XML format is under development.

Nearly 1,300 individuals and organizations license the UMLS Knowledge Sources and associated programs. The UMLS data are available in relational format via file transfer protocol, on CD-ROMs, or through the API or interactive use of the UMLS Knowledge Source server. Responsibility for distribution of UMLS data was transferred from the Office of the Associate Director, Library Operations to MMS during FY 2000. The Lister Hill Center and LO are working together to develop additional helpful information for UMLS licensees for NLM's web site.

Print and Electronic Publications

NLM publishes some of its authoritative data in print publications, including *Index Medicus*, the *List of Journals Indexed in Index Medicus*, and several MeSH publications, but regards its electronic databases as the primary means of making these data available. In recognition of this fact, NLM signed a formal agreement with the Government Printing Office that it will continue to provide public access to current and historical indexing and cataloging data. This agreement relieves regional depository libraries of the necessity to retain retrospective volumes of *Index Medicus* and the *NLM Current Catalog*. Of course NLM will continue to retain the printed volumes in its collection.

The Library continues to review and modify or eliminate specific print publications that have outlived their usefulness, given increasing user access to more flexible forms of NLM data. As NLM replaces its legacy systems, the programs that produce any print publications

which the Library intends to continue must also be replaced. In FY2000, LO and OCCS designed and implemented a new method of producing the *List of Journals Indexed in Index Medicus* from the Voyager ILS. LO decided to cease printing the *List of Serials Indexed for Online Users*, but made it available in PDF format for local printing. At least one commercial firm decided to produce and sell printed copies from the PDF version.

The NLM World Wide Web site has become the primary vehicle for distributing a wide range of publications from fact sheets to technical reports to multimedia catalogs. PSD serves as the manager for NLM's main web site. There were 2.1 million hits to its publication pages in FY2000, a 5% increase from the previous year. Among the most popular publications are issues in the *Current Bibliographies in Medicine* series, which is edited by the Reference and Customer Services Section. Each bibliography addresses a topic of current interest to NLM, NIH, or other federal agencies and may be produced in conjunction with an NIH consensus development conference, a White House conference, or another meeting. Often the topics are difficult to search in NLM's databases or are spread across the literature of multiple disciplines. Reference and sometimes NICHSR staff members collaborate with outside experts to produce each bibliography. In FY2000, the practice of linking citations in the bibliographies to corresponding MEDLINE citations in PubMed was initiated, and LO began publishing the bibliographies in PDF as well as HTML format. FY2000 additions to the series included: Adjuvant Therapy for Breast Cancer; Visible Human Project; Phenylketonuria (PKU): Screening and Management; National Nutrition Summit: Information Resources; Osteoporosis; Health Literacy; Improving Implant Performance through Retrieval Information; and Bioavailability of Nutrients and Other Active Components of Dietary Supplements.

In an illustration of what can be achieved via web publishing, HMD worked with Emilie Savage-Smith, Ph.D., an eminent scholar of Islamic medical history, to produce a catalog *raisonne* of *Islamic Medical Manuscripts at the National Library of Medicine*, which was

released in May 2000. This beautiful catalog makes effective use of the technology to link catalog records, descriptive and evaluative text, calligraphy, and full-color images of pages from the early manuscripts with supporting biographical sketches and definitions of key terms. It would have been prohibitively expensive to produce a less flexible version of this work in printed form.

Direct User Services

In addition to its electronic and printed products, LO provides document delivery, reference and customer service as a national and international backup to services available from other health sciences libraries and information suppliers. LO also serves a large onsite clientele in the NLM reading rooms.

Document Delivery

LO provides copies of documents in the NLM collection to other U.S. and international libraries to fill requests from health professionals, researchers, and other interested people which cannot be readily be filled by other members of the National Network of Libraries of Medicine, libraries in other countries, or other document suppliers. LO also retrieves documents from the Library's closed stacks for use by onsite patrons.

In FY2000, PSD's Collection Access Section processed a total of 749,373 document requests, essentially equal to the number received last year. Onsite users requested 359,295 documents from NLM's closed stacks, one percent more than in FY1999. Remote libraries submitted 390,574 interlibrary loan (ILL) requests, about 1.5% less than last year. NLM filled 36% of its ILL requests electronically.

The number of ILL requests received by NLM exceeded the FY1999 level for the first three quarters of the fiscal year, but then declined 20% after the implementation of the new DOCLINE system in July 2000. The new DOCLINE system provides network libraries with additional options for routing ILL requests, including the ability to route requests to any NN/LM resource library that holds the material

nationwide before the request is sent on to NLM. The new routing features are probably responsible for the decline in requests routed to NLM. Fortunately, the drop in NLM ILL traffic does not appear to have resulted in unwanted increases in demand for other NN/LM libraries. The 3,187 libraries that are DOCLINE users entered a total of 2.985 million requests into the system in FY2000, about 1% less than in FY1999. Ninety-five percent of the requests were filled.

A major component of System Reinvention, the new web-based DOCLINE system combines the functionality available in the previously separate DOCLINE, SERHOLD, and DOCUSER systems; interfaces with the PubMed and LOCATOR^{plus} retrieval systems; and supports more flexible ILL request generation and online updating of routing tables and serial holdings data by any network member. Because the original DOCLINE system provided unique request generation and automated routing functionality, the development of its replacement required significant custom development by OCCS working with staff from PSD, the Serial Records Section, the NN/LM office, and the Regional Medical Libraries (RMLs). DOCLINE is central to resource sharing within the NN/LM, and more than 2,000 network members had to learn to use the new system and integrate it into their local operations. To facilitate this process, PSD created a special listserv, DOCLINE-L, and a special web site for important announcements about the new DOCLINE. The RMLs created many special training materials and provided numerous training sessions for network members. At the end of FY2000, the basic functionality of the new DOCLINE was working well, but several important tasks remained to be done, including completion of the routines for producing document delivery statistics and NLM's interlibrary loan bills, a number of enhancements requested by the RMLs, and implementation of the ISO ILL protocol to facilitate exchange of ILL requests between DOCLINE and other systems. Implementation of the protocol is complicated by the fact that each vendor system applies its own set of communication requirements. NLM is testing

with three external systems and expects to implement the protocol fully in FY 2001.

Loansome Doc is a system that allows individuals to automatically route requests for documents identified in MEDLINE to a specific library that has agreed to serve them. During FY2000, PSD rewrote and simplified the Loansome Doc registration and ordering information that tells individual PubMed and Internet Grateful Med searchers how they may register with a library to request documents. Users requested 820,777 documents via Loansome Doc in FY2000, a greater than 80% increase from the previous year. PSD also worked with a variety of U.S. and international libraries to expand access to Loansome Doc and DOCLINE for individual users and libraries worldwide. Thirty-five libraries outside the U.S. and Canada now use DOCLINE and 13 of them provide Loansome Doc service. In a special international document delivery project that is part of the Multilateral Initiative on Malaria, staff from NICHSR and PSD worked with NLM's Office of Health Information Programs Development to arrange for document delivery to several Malaria Research Centers in Africa.

Reference and Customer Service

The Public Services Division and the History of Medicine Division provide reference and research assistance to onsite and remote users as a backup to services available from other health science libraries. PSD's Reference and Customer Service Section also has primary responsibility for responding to inquiries from those seeking information about NLM products and services or assistance in using these services. Staff throughout LO and NLM provide second-level service for questions that cannot be answered by first-line customer service staff.

In FY2000, the Reference and Customer Service Section handled 114,427 user requests, a 2.8% increase from the previous year. Offsite requests increased 15% to 62,971. First-line customer service staff handled 80% of the inquiries. During the past year, the customer service staff assumed responsibility for responding to inquiries about ClinicalTrials.gov and document delivery, including the new DOCLINE system. The number of offsite

inquiries received via e-mail increased 33% to 47,924; the number of requests received via telephone decreased 19% to 14,762. Onsite inquiries decreased 9% to 51,456, as Reading Room patrons became more familiar with the new online catalog and circulation system implemented last year.

Work continues on the creation and revisions of answers to "Frequently Asked Questions" for NLM's public web sites and stock replies for use by customer service staff. The Reference and Customer Service Section also developed a web-based tutorial on "How to research a medical topic" for use by students and other members of the general public.

In June 2000, the Reference Section conducted a customer satisfaction survey of Reading Room patrons. Eighty-seven percent of the patrons surveyed responded that they visited NLM to "find materials on a specific topic" or "find a specific book, journal, or audiovisual." Ninety-six percent were able to "totally or partially" find what they seeking and 98% reported that the quality of service was excellent, good, or satisfactory. PSD and HMD provide reference and research assistance to onsite and remote users as a backup to services available from other health sciences libraries.

Outreach

Many LO programs are designed to increase awareness and use of NLM's services by librarians and other information providers, health professionals, researchers, and the general public. LO coordinates the National Network of Libraries of Medicine (NN/LM) which attempts to equalize access to health information services and technology for health sciences libraries, health professionals, and the general public throughout the United States; participates in NLM-wide efforts to develop and evaluate outreach programs designed to improve health information access for underserved minorities and the general public; develops major exhibitions and other special programs in the history of medicine; and conducts a range of training programs for health sciences librarians. Many LO staff members give presentations and demonstrations at professional meetings and

write articles to highlight NLM programs and services.

National Network of Libraries of Medicine

The goal of the NN/LM is to provide U.S. health professionals, researchers, educators, administrators, and members of the public with timely, convenient access to biomedical and health information resources. The NN/LM strives to ensure that accurate and current information is available irrespective of the user's geographic location or institutional affiliation. The network has more than 4,500 health sciences libraries, including hospital and academic medical center libraries, located throughout country. LO's NN/LM Office oversees the network programs that are coordinated and administered by eight Regional Medical Libraries (RMLs) under contract to NLM. (See Appendix 1 for a list of the RMLs.) As mentioned previously, the 5-year RML contracts are currently being re-competed. The statement of work for the 2001–2006 contracts adds an emphasis on outreach to the general public while continuing to focus on coordination and support for network members and outreach to health professionals, particularly those serving minority groups and working in rural areas and inner cities.

The NN/LM is a core component of NLM's outreach program and its efforts to reduce health disparities. The RMLs and other NN/LM members develop and conduct many special projects to reach underserved health care professionals and to improve the public's access to high quality health information. In FY 2000, the Library funded four special NN/LM outreach projects designed to improve health professionals access to information services and 62 projects to increase the public's access to high quality health information. Many of these projects involve partnerships between health sciences libraries and other organizations, including public health departments, professional associations, public libraries, schools, and community-based organizations. Examples:

- The Tifton-Tift Public Library is reaching out to 75% of the rural Southern Georgia community to introduce online health

information in sessions at non-traditional settings such as churches, service centers, and health clinics as well as at libraries.

- The University of North Carolina at Chapel Hill Health Sciences Library and the School of Information and Library Science are partnering to improve online public access to national and local health information resources. They are assisting NLM in determining how data from MEDLINEplus can best be integrated with descriptions of locally applicable resources.
- The Spencer S. Eccles Health Sciences Library and the Office of Patient Education at the University of Utah Hospitals and Clinics are translating patient education materials into Spanish and creating a bilingual search engine to afford easy access to health materials for the Hispanic population of Utah.
- "EMPOWERMENTplus" is a collaborative project between Healthnet: Connecticut Consumer Health Information Network and the Connecticut Self-help Network designed to train representatives from self-help groups to access and evaluate online healthcare information and to train other group members.
- The PARTNERS (Primary Care Access to Resources, Training, Networks, Education and Research Services) project provides information technology resources and training to 10 non-profit community-based clinics in Washington, D.C.
- A project of the Louisiana State University Health Sciences Center Library, Shreveport, will extend access to electronic health data to the heavily concentrated minority populations of rural parishes of the lower Mississippi Delta and to those in under-served areas of Northwestern Louisiana.

In FY2000, the Region 6 RML at the University of Washington completed work on a joint NLM/RML project to develop *Measuring the Difference: Guide to Planning and Evaluating Health Information Outreach*, which will be generally applied throughout the NN/LM. The Region 6 RML is funded by NLM

to serve as a consultant on outreach evaluation for the entire network.

The NN/LM is currently focusing on underserved public health professionals as part of its participation in Partners in Information Access for Public Health Professionals, a joint effort of NLM, the NN/LM, the Centers for Disease Control and Prevention, the Health Resources and Services Administration, the Association of State and Territorial Health Officials, the National Association of County and City Health Officials, and the Public Health Foundation. The NN/LM Office, NICHSR, and the Specialized Information Services Division coordinate NLM's participation in this partnership, which is designed to improve access to advanced information technology and information services for practicing public health professionals. As part of this initiative, NICHSR has arranged NLM and NN/LM viewing sites for CDC-sponsored Public Health Grand Rounds programs, represents NLM on the Public Health Workforce Development Collaborative, and has provided funding and technical advice to the Public Health Foundation for efforts to develop web-based information resources targeted toward assisting public health officials in addressing Healthy People 2010 objectives. NICHSR is also working with NN/LM members to organize a 2001 conference to review NN/LM public health outreach efforts to date and to determine what has been learned that should be applied to future initiatives.

The RMLs and other NN/LM members conduct most of the exhibits and demonstrations of NLM products and services at health professional, consumer health, and general library association meetings around the country. LO staffs exhibits at the annual meeting of the Medical Library Association, some of the health professional and library meetings held in the Washington, D.C. area, and some distant meetings focused on health services research, public health, and the history of medicine. In FY2000, NLM and NN/LM services were displayed at 177 exhibits at national, regional, and state association meetings across the U.S. There was increased emphasis on meetings of public and school librarians and consumer groups. BSD created posters, bookmarks, and other materials that highlight MEDLINE*plus*,

ClinicalTrials.gov, and other NLM services that are particularly useful to the general public and the public and school librarians who serve them.

Special NLM Outreach Initiatives

LO contributes to a number of NLM-wide efforts to expand outreach and services to the general public and to address racial and ethnic health disparities. In FY2000, staff members from the Office of the Associate Director, the NN/LM Office, PSD, and BSD were active participants in the NLM-wide Consumer Health Coordinating Committee chaired by NLM's Assistant Director for Research and Education. Staff from the Office of the Associate Director and PSD also play key roles in the NLM-wide Web Evaluation Committee, chaired by the Associate Director for Health Information Programs Development. The purpose of this Committee is to develop improved methods for measuring consumer use of NLM's web-based services and evaluating the quality and impact of these services.

In addition to the development of MEDLINE*plus* and the NN/LM outreach activities described elsewhere in this report, LO staff members initiated planning for a joint NLM/Public Libraries Association/Medical Library Association conference on the "The Public Library and Consumer Health" to be held in 2001 in conjunction with the American Library Association's midwinter meeting. LO also made arrangements to mail information about MEDLINE*plus* and other NLM services to all public and NN/LM libraries in eight states. LO plays a key role in the Library's "Adopt-a-School" partnership with Wilson High School in the District of Columbia, helping to organize access to health information in the school and providing related training. LO also directs an outreach program targeted toward community-based organizations that serve minority populations, with a goal of helping these groups to compete effectively for NLM funding opportunities. In FY2000, this program led to an increase in applications from, and AIDS outreach awards to, such community organizations.

LO assisted the NLM Office of Administration in applying successfully to the

Office of Management and Budget for blanket authority to conduct customer surveys. This authority substantially reduces the lead time required for approval of such surveys. LO also chairs the subgroup of the Web Evaluation Committee that is developing an online survey instrument for MEDLINEplus users. This instrument is being designed and tested so that it can be easily be modified for use in evaluation of other NLM web-based services. LO is working with Regions 6 and 7 in the NN/LM to test the use of semi-structured interviews with public library users as a method for obtaining information on the usefulness and impact of NLM's health information services.

Historical Exhibitions and Programs

HMD periodically mounts major exhibitions in the NLM lobby and rotunda, with assistance from the Lister Hill Center, the Office of Communications and Public Liaison, the Office of Administration, and the Office of the Director. Designed for the interested public as well as the specialist, these exhibitions are part of NLM's outreach program. *Breath of Life*, the current exhibition on the history of asthma and the state of knowledge about the disease, was developed by HMD in collaboration with the National Heart, Lung, and Blood Institute, the National Institute of Allergy and Infectious Diseases, and the National Institute of Environmental Health Sciences. During FY2000, the Lister Hill Center developed a Digital Video Disk traveling version of this exhibition in response to a request from the Office of the Surgeon General, which had a successful debut at a major medical meeting in June 2000. HMD devoted substantial time and resources to the development of the next major exhibition, *The Once and Future Web*, which will open at NLM in May 2001. It will examine parallels between the telegraph system and the World Wide Web and current and future implications of the web for medicine and health.

In May 2000, HMD mounted a special exhibit of about 25 medieval manuscripts from its collection in the HMD Reading Room. *Art is long, Life is short* commemorated the return to NLM in FY1999 of the Latin manuscript, "Treatises on Medicine," written in England in

the 12th century on vellum (calf skin), which had disappeared from the Library some 50 years ago. The exhibit opened in conjunction with a reception held at the NLM for attendees at the annual meeting of the American Association of the History of Medicine. A number of NLM staff members donned medieval costumes for this event, which also featured medieval music.

HMD routinely installs "mini-exhibits" in the exhibit cases at the entrance to the HMD Reading Room. At the start of FY2000, an exhibit of *Classics of Traditional Chinese Medicine* was on display in these cases. It was replaced in July 2000 by *Joshua Lederberg: Biomedical Science and the Public Interest*, an exhibit highlighting the career of the Nobel prize winning scientist in conjunction with his 75th birthday. Dr. Lederberg is a member of the NLM Board of Regents and chairs the PubMed Central Advisory Committee. His papers are also featured in *Profiles in Science*.

As part of its broader initiative to increase the amount of information related to the history of medicine available on the web (see other sections of this report), HMD creates online web versions of many of exhibitions and smaller exhibits. In FY2000, the online versions of *Emotions and Disease* and "*That Girl There is Doctor of Medicine*": *Elizabeth Blackwell, America's First Woman M.D* became available. HMD also mounted a web tour of *Historic Medical Sites in the Washington, D.C. Area*.

HMD sponsors a series of seminars by historical scholars as well as special public lectures in cooperation with the NLM Diversity Council. Vickie M. Mays, Ph.D. presented the African American History lecture "Racism, Sexism, and Poverty are Hazardous to Our Health" on March 16. On April 4, Sandra Long, Ph.D. presented a lecture on "Women in Science and Medicine: What Difference Does It Make?" Bert Hansen, Ph.D. lectured on "Has the Laboratory Been a Closet? Gay and Lesbian Lives in the History of Science" on June 15.

NICHSR and the Lister Hill Center worked with HMD to complete *Health Services Research: A Historical Perspective*, a video history of the field for use in NLM training courses for health sciences librarians and in academic health services research programs. Theodore Brown, Ph.D. wrote the script for the

well-received video, which makes use of material from oral and video history interviews with key figures commissioned by NICHSR. In January 2001, NICHSR will convene an ad hoc advisory committee of health services researchers, historians, and librarians to advise NLM on appropriate next steps in its initiative to document and preserve the history of health services research.

HMD staff members presented historical papers and lectures at professional meetings throughout the year and also published the results of their scholarship in books, chapters, articles, and reviews. In FY2000, HMD joined forces with William Helfand, a long-time benefactor of NLM's historical collections, to initiate a monthly "Images from the History of Public Health" feature in the *American Journal of Public Health*. This series often features pictures from NLM's collection.

Simon Baatz, Ph.D. joined the NLM staff as a temporary employee in May 2000 to undertake a history of NLM focusing on developments in the last 20 years. HMD has established a program to bring other historians to NLM for 1-3 months to assess and use the NLM collections. Walter Lear, M.D., an authority on social medicine, came to the Library under this program in FY2000 and produced reports and recommendations regarding NLM's priorities for collecting manuscript collections.

Training and Recruitment Programs for Health Sciences Librarians

LO develops online services training programs for health sciences librarians and other search intermediaries; oversees the activities of the NN/LM-funded National Online Training Center at the New York Academy of Medicine; directs the NLM Associate Fellowship program for post-masters librarians; and develops and presents continuing education programs for librarians in health services research, public health, and related topics. LO also collaborates with the Medical Library Association and other relevant groups to increase the diversity of those pursuing careers in health sciences librarianship.

In FY 2000, MMS and the National Online Training Center taught a total of 1,118 students in 80 separate classes. Training manuals

for the online training classes are available in PDF format on the National Online Training Center web site. The alpha version of web-based interactive courseware for PubMed training became available for testing in September 2000.

In June 2000, the four initial participants in the new optional second year of the Associate Fellowship program returned to NLM with their preceptors to discuss their experiences and provide advice to the Library on how to improve the program. In the second year, Associates work on a multidisciplinary team engaged in developing information systems or services relevant to the clinical, educational, or research missions of the hosting institutions. Both the Associates and their host sites were enthusiastic about the program and offered good suggestions for minor enhancements. Again in FY2000, four first-year Associates elected to participate in the second year at McMaster University, the University of Pittsburgh, Johns Hopkins University, and Vanderbilt University. The three other Associates accepted positions at Harvard University's Digital Library program, with the Kevric Company providing customer support to users of NCBI's genetic databases, and with Science Applications International Incorporated (SAIC), providing web site support to the National Institute of Science and Technology. Seven Associate Fellows began the first year of the program at NLM in September 2000; two are members of minority groups. An international Associate Fellow from the Republic of China will join the program early in FY2001.

NICHSR continues to develop continuing education programs to increase health sciences librarians' understanding of health services research and related fields. In addition to the video history described in the last section, NICHSR commissioned the development of a course on *Finding and Using Health Statistics*, which was first presented at the annual meeting of the Medical Library Association in May 2000. The course was subsequently mounted on the NICHSR web site as a self-study course (www.nlm.nih.gov/nichsr/usestats/index.htm). NICHSR is currently working on a symposium on "Library Partnerships: Making Powerful Connections" which has been selected for

presentation in conjunction with the 2001 MLA Meeting.

In FY2000 LO worked through Region 5 Regional Medical Library at the Houston Academy of Medicine/Texas Medical Center Library to provide funds to the Medical Library Association to encourage minority students to choose health science librarianship as a career. The support will enable MLA to strengthen its programs for recruiting minorities into the medical library profession and to increase scholarship opportunities for minority students seeking degrees in librarianship. NLM funds will be used to increase the size of the MLA's existing minority scholarship, to support, in partnership with MLA, the American Library Association's Spectrum Scholars program to attract students of color to graduate programs in library and information studies, and for outreach to minority college and high school students. LO continues to create special web pages highlighting important projects undertaken by health sciences librarians during October, which MLA has designated as National Medical Librarians Month.

The 2000 NLM/MLA Joseph Leiter Lecture was held at NLM on May 17, 2000, in conjunction with an NLM Board of Regents meeting. Dr. Scott C. Ratzan, Editor-in-Chief, *Journal of Health Communication*, discussed "Quality Communication: The Path to Ideal Health."

Health Informatics Activities

In addition to providing the Library's basic services, LO represents NLM in several initiatives designed to promote more effective health applications of advanced computing and communications technologies. In FY2000, LO continued to serve on the Department's Health Data Standards Committee that is overseeing the implementation of the administrative simplification provisions of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). LO assisted in drafting the language related to codes and classifications that appeared in the final HIPAA Transactions and Code Sets regulation, which was published on August 17, 2000. In FY1999, LO initiated and now directs a contract jointly funded by HHS, the Department

of Defense, and the Department of Veterans Affairs that supports the continued development and free distribution of LOINC (Logical Observations: Identifiers, Names, Codes), a detailed clinical coding system that will be part of the HIPAA claims attachments standard. In FY2000 and again on behalf of other Federal agencies, LO began negotiations with the College of American Pathologists for a sole source contract that would provide a broad U.S. license for use of SNOMED (The Systematized Nomenclature of Medicine) in health data systems. If the negotiations are successful, the contract will be awarded in a future fiscal year.

NICHSR worked with NLM's Extramural Programs Division to arrange a joint NLM/Agency for Healthcare Research and Quality (AHRQ) workshop on "Medical Informatics and Health Services Research: Bridging the Gap," which was held at NLM on January 6-7, 2000. The goal of the workshop was to discuss the need for additional investigators trained to work at the intersection of informatics and health services research and to recommend ways to expand the pool of such investigators. A summary of the meeting and a preliminary set of recommendations has been published on the web (www.nlm.nih.gov/nichsr/mihsr/mihsrrec.html). A series of papers is being readied for publication. NICHSR is also working on a sequel to the issue of Current Bibliographies in Medicine on Public Health Informatics published in 1996. The new bibliography is being prepared for use in conjunction with the planned 2001 Spring Congress of the American Medical Informatics Association on public health informatics. LO is assisting with planning the meeting.

NLM was one of several funders of a workshop on "Improving Access to and Confidentiality of Research Data" convened by the Committee on National Statistics of the National Research Council in October 1999. The report of the workshop was published in 2000.

In FY2000, several LO staff members continued to serve as project officers on telemedicine evaluation contracts supported by NLM's Office of High Performance Computing and Communications.

Table 1**Growth of Collections**

<i>Collection</i>	<i>Previous Total (9/30/99)</i>	<i>Added FY 2000</i>	<i>New Total (9/30/00)</i>
<i>Book Materials</i>			
<i>Monographs:</i>			
Before 1500	578	0	578
1501-1600	5,814	4	5818
1601-1700	10,133	6	10,139
1701-1800	24,477	6	24,483
1801-1870	41,155	13	41,168
Americana	2,341	0	2,341
1870-Present	668,014	14,136	682,150
Theses (historical)	281,794	0	281,794
Pamphlets	172,021	0	172,021
Bound serial volumes	1,156,513	33,868	1,190,381
Volumes withdrawn	(68,735)	(3,825)	(72,560)
Total volumes	2,294,105	44,208	2,338,313
<i>Nonbook Materials</i>			
<i>Microforms:</i>			
Reels of microfilm	102,760	6,248	109,008
Number of microfiche	406,422	14,009	420,431
Total microforms	509,182	20,257	529,439
Audiovisuals	64,933	1,543	66,476
Computer software	1,520	260	1,780
Pictures	56,684	256	56,940
Manuscripts	2,858,957	87,150	2,946,107
Total nonbook	3,491,276	109,466	3,600,742
Total book and nonbook	5,785,381	153,674	5,939,055

Table 2**Acquisition Statistics**

<i>Acquisitions</i>	<i>FY 1998</i>	<i>FY 1999</i>	<i>FY 2000</i>
Serial titles received	22,247	22,433	23,141
Publications processed:			
Serial pieces	146,921	123,823	143,636
Other	21,642	14,418	22,384
Total	168,563	138,241	166,020
Obligations for:			
Publications	\$5,266,996	\$5,370,797	\$4,895,999
(For rare books)	(\$251,293)	(\$292,603)	(\$267,300)

Table 3**Cataloging Statistics**

	<i>FY 1999</i>	<i>FY 1999</i>	<i>FY 2000</i>
Completed Cataloging.....	18,803.....	14,396.....	20,067

Table 4**Bibliographic Services**

<i>Services</i>	<i>FY 1998</i>	<i>FY 1999</i>	<i>FY 2000</i>
Citations published in MEDLINE.....	411,921.....	434,525.....	442,168
For <i>Index Medicus</i> *.....	388,022.....	421,423.....	434,813
Journals indexed for <i>Index Medicus</i>	3,302.....	3,394.....	3,472
Abstracts entered.....	312,064.....	338,435.....	341,682

Table 5**Circulation Statistics**

<i>Activity</i>	<i>FY 1998</i>	<i>FY 1999</i>	<i>FY 2000</i>
Requests Received.....	694,281.....	751,732.....	749,869
Interlibrary Loan.....	374,791.....	396,516.....	390,574
Onsite.....	319,490.....	355,216.....	359,295
Requests Filled:.....	523,081.....	570,966.....	589,516
Interlibrary Loan.....	275,588.....	301,073.....	299,182
Photocopy.....	264,301.....	291,743.....	290,472
Original.....	10,167.....	8,229.....	8,710*
Onsite.....	247,493.....	269,893.....	292,664

*Beginning in FY 2000 "original" includes audiovisual materials loaned.

Table 6

Online Searches—All Databases

	<i>FY 1998</i>	<i>FY 1999</i>	<i>FY 2000</i>
Total online searches.....	104,000,000.....	191,000,000.....	244,000,000

Table 7

Reference and Customer Services

<i>Activity FY 1997</i>	<i>FY 1998</i>	<i>FY 1999</i>	<i>FY 2000</i>
Offsite requests.....	27,070.....	54,542.....	62,971
Onsite requests.....	43,782.....	56,737.....	51,456
Total.....	70,852.....	111,279.....	114,427

Table 8

Preservation Activities

<i>Activity</i>	<i>FY 2000</i>
Volumes bound.....	31,874
Volumes microfilmed.....	4,513
Volumes repaired onsite.....	2,000
Audiovisuals preserved.....	46
Historical volumes conserved.....	385

Table 9

History of Medicine Activities

<i>Activity FY 1997</i>	<i>FY 1998</i>	<i>FY 1999</i>	<i>FY 2000</i>
Acquisitions:			
Books.....	108.....	170.....	226
Modern manuscripts.....	274,530.....	129,885.....	1,915,550
Prints and photographs.....	849.....	1,773.....	1,391
Historical audiovisuals.....	94.....	114.....	37
Processing:			
Books cataloged.....	193.....	58.....	49
Modern manuscripts cataloged.....	0.....	0.....	87,150
Pictures cataloged.....	0.....	83.....	256
Citations indexed.....	1,516.....	1,022.....	1,066
Public Services:			
Reference questions answered.....	12,387.....	14,050.....	15,143
Onsite requests filled.....	3,733.....	3,672.....	4,485

SPECIALIZED INFORMATION SERVICES

*Steven J. Phillips, M.D.
Acting Associate Director*

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established more than 30 years ago at NLM in the Division of Specialized Information Services (SIS). Over the years TEHIP has evolved to provide for the increasing need for toxicological and environmental health information, taking advantage of new computer and communication technologies to provide more rapid access to a wider audience. Our development of novel search capabilities means that users need not have extensive knowledge of searching techniques and thus allows data to be relayed to them more effectively. Finally, we have moved beyond the bounds of the physical National Library of Medicine, exploring ways to point and link users to relevant sources of toxicological and environmental health information wherever these sources may reside. This is being accomplished primarily through the TEHIP and AIDS web sites developed and maintained by SIS. Development of HIV/AIDS information resources became a focus of the Division several years ago, and now includes several collaborative efforts in information resource development and deployment. Continuous refinements and additions to our web-based systems are made to allow easy access to the wide range of information collected by this Division. Our usage has continued to increase over the past year with access to all toxicology and HIV/AIDS data free over the Internet.

In FY 2000 SIS reexamined the scope and coverage of current programs, selecting several for significant re-engineering. We proposed new opportunities to enhance SIS information services and provide new services in emerging areas. This examination has been guided in the past by two Institute of Medicine reports focusing on the TEHIP Program: *Toxicology and Environmental Health*

Information Resources: the Role of the National Library of Medicine, released in the spring of 1997, and a follow-on report, *Internet Access to the NLM's Toxicology and Environmental Health Databases*, published in 1999. Both reports have been instrumental in our re-engineering efforts, and were used as starting points for internal staff discussions at a strategic planning retreat held in April 2000.

Resource Building

The wide range of resources related to toxicology and environmental health information and HIV/AIDS information include many databases that are created or acquired as well as other services and projects.

The **Hazardous Substances Data Bank** (HSDB[®]) continues to be a highly used resource, averaging over 25,000 searches each month. Increased emphasis continues to be placed on providing more data on human toxicology and clinical medicine within HSDB, in keeping with past recommendations of the Board of Regents' Subcommittee on TEHIP. The selection of new members of the Scientific Review Panel for HSDB reflects this shift in content emphasis. Newer sources of relevant data are being examined for incorporation into new and existing data fields within the current 4,550 HSDB records. Because of increased staff efforts, more records are being processed through special enhancements, including source updates from various peer-reviewed files. The process of developing a new web-based system for HSDB creation, review, and maintenance has begun. An initial workshop to define some of the issues related to this re-engineering effort will be held in October, 2000.

ChemID[®] (Chemical Identification File) is an NLM online chemical dictionary that contains over 350,000 records, primarily describing chemicals of biomedical and regulatory importance. It is available to users through Internet Grateful Med, and also on the web as the ChemIDplus file. ChemIDplus has additional features, including chemical structure search and display for 68,000 chemicals, and hyperlinked locators that retrieve data for a given chemical from other resources such as MEDLINE or HSDB. Over 15,000 records of

regulatory interest collectively known as SUPERLIST are also available and hyperlinked in ChemIDplus. During FY2000, an online web maintenance system was developed for ChemIDplus, allowing individual record correction and addition to this resource. A prototype batch system was also developed to allow multiple record corrections. Chemical structure addition and maintenance is done in a method that also allows immediate updating. Over 12,000 structures were added to ChemIDplus in FY2000.

TOXLINE[®] (Toxicology Information Online) is a large bibliographic database traditionally produced by merging "toxicology" subsets from some 18 secondary sources. By the end of FY2000, the database included nearly 3 million citations to toxicology literature going back to 1965. In FY2000, we began the transition to a next generation TOXLINE, reducing the components needed to produce the database by creating a toxicology subset on NLM's PubMed so that users can access standard journal literature in toxicology and environmental health as part of an enlarging MEDLINE database. We are adding more journals in the area of toxicology and environmental health to MEDLINE to cover some of the literature formerly provided by outside sources. For the nonstandard journal literature in this area we are creating a web-based system on TOXNET that will allow efficient acquisition and updating of these components. The next generation TOXLINE will be available to users on distributed systems, with an integrated approach provided by the new NLM Gateway and new features of the TOXNET[®] search system.

DIRLINE[®] (Directory of Information Resources Online) is NLM's online directory of resources of organizations, databases, bulletin boards, as well as projects and programs with special biomedical subject focus. These resources provide information to users which may not be available from one of the other NLM bibliographic or factual databases. DIRLINE continues to receive a high level of use through a new interface that became public in October 1999. This new interface supports direct links to the web sites of the organizations listed in the database, as well as direct e-mail connections.

The quality and utility of the database continues, for example, duplicates have been eliminated. *Health Hotlines*, the always popular publication of health-related toll-free telephone numbers, has a web version that also indicates the availability of Spanish-speaking customer service representatives and Spanish language publications from the resources listed.

The **Toxic Chemical Release Inventory** (TRI) series of files now includes four online files, TRI95 through TRI98. These files remain an important resource for environmental release data and are a useful complement to our other databases. Mandated by the Emergency Planning and Community Right-to-Know Act (Title III of the Superfund Amendments and Reauthorization Act of 1986), these EPA databases contain data on environmental releases to air, water, and soil for over 600 EPA-specified chemicals.

The **Chemical Carcinogenesis Research Information System** (CCRIS) continues to be built, maintained, and made publicly accessible at NLM. This data bank is supported by the National Cancer Institute (NCI) and has grown to over 8,000 records. The chemical-specific data covers the areas of carcinogenesis, mutagenesis, tumor promotion and tumor inhibition.

The **Integrated Risk Information System** (IRIS), EPA's official health risk assessment file, continues to experience high usage and be very popular. EPA has had a version of IRIS on the agency's web page since 1996, and as we move to web access we will consider how best to integrate our web service with what EPA provides. IRIS now contains 535 chemicals.

The **GENE-TOX** file continues to be built and updated directly on TOXNET by EPA scientific staff. This file contains peer-reviewed genetic toxicology (mutagenicity) studies for about 3,200 chemicals. GENE-TOX is popular with users in other countries.

The **Registry of Toxic Effects of Chemical Substances** (RTECS) is a data bank based upon a National Institute for Occupational Safety and Health (NIOSH) file by the same name which NLM restructured and made available for online searching. With our move to free Internet access to all databases, NIOSH

requested that we no longer include RTECS on our system. We continue to use RTECS in the creation of the Hazardous Substance Data Bank.

The **Developmental and Reproductive Toxicology** (DART[®]) database now contains over 46,000 citations from literature published since 1989 on agents that may cause birth defects. DART is a continuation of the Environmental Teratology Information Center backfile (ETICBACK) database, which contains almost 50,000 citations to literature published from 1950 to 1989. DART is funded by NLM, the EPA, the National Institute of Environmental Health Sciences (NIEHS) and the FDA's National Center for Toxicological Research and is managed by NLM.

The **Environmental Mutagen Information Center** (EMIC) database contains over 24,000 citations to literature on agents that have been tested for genotoxic activity. A backfile for EMIC (EMICBACK) contains over 75,000 citations to the literature published from 1950 to 1991. The EPA, NIEHS, and NLM, collaborating partners in this effort, decided to stop compiling this special collection as of December, 1999.

Resource Access

The SIS web server provides a central point of access for the varied programs, activities, and services of the Division. Through this server (sis.nlm.nih.gov) users can access interactive retrieval services in toxicology and environmental health or HIV/AIDS information, find program descriptions and documentation, or be connected to outside related resources. Both the toxicology and environmental health and AIDS web pages provide links to NLM outreach activities in these subjects, access to NLM databases, links to selected web sites, as well as tutorials, fact sheets, and other publications produced by SIS. Over 8,000 users visit the SIS web site weekly and view approximately 50,000 pages.

Toxicology Data Network (TOXNET)

The **Toxicology Data Network** (TOXNET), NLM's computer system providing data bank building for many of its toxicology

files, has moved from a networked microprocessor environment to a UNIX-based platform (Solaris Version 2.6) on a SUN Enterprise 3000 computer. Integration of this configuration with other SIS database creation systems and the web access to them is currently under way.

In FY2000, SIS continued to develop the new search interface to access all of the SIS toxicology and environmental health databases. This new search interface allows users to easily search HSDB, TOXLINE, CCRIS, Gene-Tox, DART, EMIC, IRIS, and TRI. Based on recommendations from the IOM, users are presented with a basic search screen with just a single input box for searching, with customized screens for more sophisticated users. These advanced features include Boolean searching and the ability to limit search terms to specific fields. By the middle of FY2000 this access became the only access for users to the TOXNET data banks, replacing an earlier Internet access. The new NLM Gateway will provide access to the TOXNET search system as well, making it easier for new users to learn about our resources.

Internet Grateful Med (IGM)

Near the end of FY1998 access to TOXLINE and ChemID was added to IGM, where access to DIRLINE, the HIV/AIDS databases, MEDLINE, and many other NLM databases was already being provided. This route of access will be discontinued during FY2001, when the ELHILL versions of TOXLINE and ChemID are terminated as part of NLM's transition from mainframe legacy systems.

Chemical Structure Server

The chemical structure server has evolved from a mechanism to provide structure searching for chemicals covered by SIS databases to a system for integrating chemical dictionary record building and structure searching. This system uses special molecular searching programs and includes a prototype database for construction of ChemID records. The chemical information resources continue to

be consolidated on a server that meets the requirements for chemical structure creation and access.

AIDS Information Services

NLM continues to refine its HIV/AIDS information services and make them more available to a wider audience. SIS staff led the development of the HIV/AIDS topic page on MEDLINE*plus*, identifying and organizing resources of specific interest to consumers. This page is a valuable addition to the NLM AIDS home page which contains information about NLM's programs, access to the HIV/AIDS-related databases, and links to selected HIV/AIDS resources of a more technical nature.

NLM has continued its successful AIDS Community Outreach Program with 16 awards in FY 2000, bringing the total number of awards made under this program to 124. This year five awards were made to enable previous recipients to expand or continue their projects. NLM-funded projects have ranged from the simple purchase of hardware and services to support a widely acclaimed web site (AEGIS), to the development of low literacy fact sheets in English and Spanish, to supporting a computer resource room in a public housing project.

At the request of our partner PHS agencies, NLM has continued its project management of the AIDS Clinical Trials Information Service (ACTIS) and the HIV/AIDS Treatment Information Service (ATIS). The ACTIS databases, AIDSTRIALS and AIDS DRUGS, are available through Internet Grateful Med, as well as on the web. The federally sponsored HIV-related treatment guidelines are also available in multiple formats on the web and in the HSTAT database.

NLM has provided training in the use of HIV/AIDS resources for different audiences. In addition to teaching the use of NLM's online resources, this training includes identifying and selecting high-quality, accurate resources. NLM works with a number of minority organizations including the National AIDS Minority Information and Education Program to provide training at regional and other meetings. In addition, NLM continues to provide training at a variety of Historically Black Colleges and

Universities (HBCUs) to faculty, staff, and members of the local community.

Outreach / User Support

SIS continues its support of the Toxicology Information Outreach Project. The objective of this initiative is to strengthen the capacity of HBCUs to train medical and other health professionals in the use of NLM's toxicological, environmental, occupational health, and hazardous wastes information resources. In addition to providing workstations, training, and free online access to HBCUs participating in the project, NLM has collaborated with the Agency for Toxic Substances and Disease Registry to train representatives from additional schools in the use of NLM's valuable online resources. This year the TIOP meeting was held in New Orleans in conjunction with the meeting of the American Association of Pharmaceutical Scientists. The meeting focused on how TIOP and its member schools could assist NLM in implementing its long-range plan.

Classes with specific user group focus have been conducted in addition to our usual NLM-based training. These include training sessions held at the annual Rural Minority Health Conference and at the Environmental Justice Resource Center at Clark Atlanta University. Another outreach effort focusing on improvement of access to health and disaster information in Nicaragua and Honduras was begun in FY2000. This project includes several components, including the development of technological infrastructure and web site enhancement.

User Support Computer-Based Activities

SIS has developed a set of Internet tutorials, *Toxicology Tutors*, which are introductory level toxicology courses available on the SIS web server. We are considering appropriate additions to this collection for development in the future. Other new avenues of user support are being focused at the consumer level, with a collaborative development of MEDLINE*plus* topics and addition of other

special topics of concern to the general public to the SIS web site.

Alternatives to Animal Testing

SIS continued to compile and publish references from the MEDLARS files that were identified as relevant to methods or procedures which could be used to reduce, refine, or replace animals in biomedical research and toxicological testing. Requests for these quarterly bibliographies have increased, as has the number of articles deemed relevant to the field. Bibliographies issued during the past four years are available on the SIS web server, and the primary distribution mechanism for this project is now the Internet.

Other Specialized Services

In addition to toxicologic data files, SIS is evaluating other areas for creating specialized

factual and bibliographic databases, for example, clinical medicine information products for public, health professional, and scientific audiences. Another area is drug information: SIS is reviewing its role in organizing and disseminating drug information in various formats and exploring whether it has a role in assessing the integrity and validity of such information. Another new project is exploring the use of a symptom- and occupation-based clinical medicine resource appropriate for use on the web. Yet another initiative is examining the utility of a web resource for consumers that links brand name household products with their ingredient chemicals and potential adverse health effects.

In these and other new initiatives, SIS continues to search for new ways to be responsive to user needs in acquiring and using toxicology and environmental health and HIV/AIDS information resources.

LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

Alexa T. McCray, Ph.D.
Director

Introduction

The Lister Hill National Center for Biomedical Communications was established by a joint resolution of the United States Congress in 1968 as a research and development division of the National Library of Medicine. Lister Hill Center research is carried out through several major programs, all sharing the purpose of improving health care information dissemination and use. Center research is conducted by drawing on a diverse set of scientific fields and methods. Researchers have backgrounds in medicine, computer science, library and information science, linguistics, engineering, and education. The Center's research activities are regularly reviewed by an outside advisory group, the Board of Scientific Counselors, whose members are drawn from the medical informatics community.

The Center is organized into five components, although many research projects involve collaboration across organizational units.

- The Audiovisual Program Development Branch conducts media development activities and supports NLM's research, development, and demonstration projects with high quality video, audio, imaging, and graphics materials. Current information about Audiovisual Program Development Branch activities appears at <http://lhncbc.nlm.nih.gov/apdb/>.
- The Cognitive Science Branch conducts research and development in computer and information science, using linguistic, statistical, and knowledge-based techniques. Current information about Cognitive Science Branch activities appears at <http://lhncbc.nlm.nih.gov/lhncbc/organization/cgsb/>.
- The Communications Engineering Branch conducts image-based research and development in such areas as document delivery, archiving, automated data entry, Internet access to biomedical multimedia databases, and imaging in support of medical educational applications. Current information about Communications Engineering Branch activities appears at <http://archive.nlm.nih.gov/>.
- The Computer Science Branch applies techniques of computer science and information science to problems in the representation, retrieval and manipulation of biomedical knowledge. Current information about Computer Science Branch activities appears at <http://lhncbc.nlm.nih.gov/lhncbc/organization/csb/>.
- The Office of High Performance Computing and Communications plans and conducts research and development activities with federal, industrial, academic, and commercial organizations concerning high performance computing initiatives. Current information about the Office of High Performance Computing and Communication activities appears at <http://lhncbc.nlm.nih.gov/lhncbc/organization/ohpcc/>.
- Lister Hill Center research activities involve both basic and applied informatics and fall into several broad research areas. Knowledge processing research includes language and information processing. Information systems research includes consumer health informatics, database systems, digital library research, and medical education systems. Image processing research includes image segmentation, compression, and transmission methods and algorithms. The Center is the focal point for NLM's high performance computing activities, including research support for telemedicine and health applications for the Next Generation

Internet. The most current information about Lister Hill Center activities can be found at <http://lhncbc.nlm.nih.gov/>.

Knowledge Processing

Unified Medical Language System

The Unified Medical Language System (UMLS) project regularly distributes a set of knowledge sources to the research community. These include the Metathesaurus, Semantic Network, and the SPECIALIST Lexicon, together with its associated lexical programs.

The Metathesaurus is a machine-readable knowledge source representing multiple biomedical vocabularies organized as concepts in a common format. It provides a rich terminology resource in which terms and vocabularies are linked by meaning. The Metathesaurus group has continued its two main tasks—producing increasingly comprehensive annual editions of the Metathesaurus with new and updated vocabulary sources, and developing and deploying new software systems for work on unified concept-oriented terminologies.

The UMLS Metathesaurus continues to grow in size, scope, and currency. There are approximately 800,000 concepts in the forthcoming 2001 release, a 10% increase since last year. The scope of the Metathesaurus is also growing to include the current and candidate DHHS standard vocabularies under HIPAA, the Health Insurance Portability and Accountability Act of 1996, in a common format and with increasing interconnections. Releases are becoming more frequent and more current. Two intra-year test releases were made this past year and quarterly releases are planned for 2001. A new automated suite of quality assurance programs and new systems to manage workflow for targeted and general editing have been developed and installed. These improvements, the port to the Oracle database management system, and documentation of the systems that create the Metathesaurus are progressing well. The goal is to allow the Metathesaurus to move from the research environment to that of a standard NLM product. The editing system is fully ported and in current use; the production system is in parallel testing. A new release

database is under development for deployment in next year. It will allow even more frequent releases of all approved concepts and will also provide update datasets.

Further research and development includes alternate, richer data structures; improvements to user tools such as the MetamorphoSys subsetting system, to assist users in selecting and using Metathesaurus data; alternative data output formats; and standard UMLS objects to provide UMLS functionality with less effort and detailed expertise.

In response to increasing needs of UMLS users for more information and help in using the UMLS, a Web site called “UMLSinfo” is being developed. When completed, it will contain added documentation, frequently asked questions, tutorials, and sample scripts. Additional user assistance is also being provided by staff from NLM’s Bibliographic Services Division.

The UMLS Knowledge Sources are made available over the Internet through the Knowledge Source Server, which provides direct access to each component of the UMLS. For example, users can request information about a particular concept in the Metathesaurus, including definition, semantic type, and synonyms as well as other concepts that are related to the input term. The Knowledge Source Server also accommodates navigation in the Semantic Network, allowing users to investigate relationships among semantic types and relations or to retrieve a list of Metathesaurus concepts assigned to a particular semantic type. Finally, the data in the SPECIALIST Lexicon is also made available, providing the user with the syntactic and morphologic information about each lexical item it contains.

The Knowledge Source Server is based on a three tier architecture. At the back end is an Oracle database that contains the UMLS data, while the middle layer consists of an application logic to handle requests from clients, either Web browsers or command line clients. There is also an API available for users who write their own applications. Most of the application logic is written in C, although a few modules are in Java. A new design is under way to re-implement the system using Java, RMI, and XML.

Currently the Knowledge Source Server is undergoing redesign and implementation based on a new UMLS object model for delivery of information to clients. The system will dynamically populate object attributes upon request to reduce transmission traffic, and multiple views of the object model will be available through a series of abstractions provided by helper methods. The new design uses Java's server registry for looking up and locating servers as well as Java DataBase Connectivity and Database Connection Pools (JDBC). A new API based on Java classes and methods for accessing object attributes will be available. In addition, there will be alternative servlets for delivery of XML encoded data to clients and a command line interface based on XML and XSL.

Medical Ontology Research

The increasing availability of online medical texts (including journal articles, encyclopedias and patient records) fuels the development of automated applications based on knowledge processing. In order to improve knowledge representation for such applications, the Lister Hill Center has recently initiated a project in medical ontology research. This project focuses on the definition, organization, visualization, and utilization of semantic spaces in the medical domain, using primarily, but not exclusively, the UMLS as its knowledge source.

Semantic spaces can be defined on the basis of semantic information provided by existing resources: medical terminologies provide relationships among concepts, while rules and facts can be drawn from knowledge bases and expert systems. Further relevant information, including lexical knowledge, co-occurrence of medical concepts, and relationships in textual databases, can be extracted from the medical literature. Additional resources such as semantic networks can be used to organize semantic spaces by helping make explicit the nature of inter-concept relationships, or, more generally, by providing an external semantic structure. Conceptual structures such as conceptual graphs can be used to represent organization in semantic spaces.

Once defined and organized, semantic spaces support visualization implementations to aid user navigation in complex knowledge representation structures. Issues such as granularity, redundancy, and consistency between sources must be addressed before designing applications for the visualization and navigation of semantic spaces. In addition to supporting visualization, semantic spaces can be utilized as the foundation for inferencing in knowledge-based systems. More generally, by specifying relationships or proximity between concepts, they provide the basic knowledge used in applications such as terminology servers as well as concept-based indexing and retrieval systems.

The Semantic Navigator, an experimental knowledge navigation tool for the UMLS, was developed as a task in this project and is now available as part as the UMLS Knowledge Source Server. Other ongoing tasks include studying ontological issues in the UMLS and defining a semantic distance between biomedical concepts.

Lexical Systems

The Lexical Systems group builds and maintains the SPECIALIST lexicon, a large syntactic lexicon of medical and general English that is released annually with the UMLS Knowledge Sources. Lexical access tools, including LVG, wordind, and norm, are also distributed with the UMLS. The lexical access services that provide information to other programs have recently been revised to provide XML-formatted records that include and identify all inflected forms.

The SPECIALIST lexicon records the spelling variation inherent in English orthography; however, it cannot deal with spelling errors. An effort is under way to investigate spelling suggestion techniques for use in terminology servers. Aspects of this multi-strategy approach have been incorporated into the spelling correction facility of the OCR component of the MARS project and the spelling correction tools attached to the ClinicalTrials.gov system. Further research in the group has resulted in a spelling correction system based on the statistics of common

misspelling patterns and fuzzy clustering around prototype words. This system improves on previous methods by providing a self-tunable, parameterized spelling suggestion algorithm.

In the past year, Lexical Systems group members implemented several modules to support the MetaMap technology transfer project. The products from this project have application as the approximate matching facility in the Knowledge Source Server; they are also used within the Indexing Initiative and can serve as a basis for further document-based IR research.

Semantic Knowledge Representation

Access to biomedical information depends on reliable representation of the knowledge contained in text. For significant advances to be achieved, a richer representation is required than is currently available. The Semantic Knowledge Representation (SKR) project develops programs that extract usable semantic information from biomedical text by building on resources currently available at NLM. The UMLS knowledge sources and the natural language processing tools provided by the SPECIALIST system are especially relevant. Two programs in particular, MetaMap and SemRep, are being evaluated, enhanced, and applied to a variety of problems in biomedical informatics. MetaMap maps noun phrases in free text to concepts in the UMLS Metathesaurus, while SemRep uses the Semantic Network to determine the relationship asserted between those concepts.

During the past year, project involvement has included research projects in biomedical information management. Most notably, MetaMap constitutes one of the core component of the Indexing Initiative system, which suggests automatically-generated indexing terms for biomedical text such as MEDLINE citations. SemRep was applied to the task of extracting semantic relationships regarding treatment and prevention from the conclusion section of structured abstracts referring to randomized clinical trials. SemRep was also used to enhance the semantic interpretation of anatomically-oriented clinical text. Syntactic constructions addressed were

those expressing the severity and specificity of location of disease.

Current research is focused in several areas aimed at enhancing the accuracy, effectiveness, and availability of SKR programs. An exportable version of MetaMap for use both internally and by the general medical informatics community is being developed. Project staff are also conducting research to resolve word sense ambiguity in natural language. The solution being sought depends on the interaction of the representation of meaning in the UMLS Metathesaurus, journal descriptor indexing, and the natural language processing tools being developed in the SKR project. Further research seeks to construct a general mechanism to accommodate the development of programs for underspecified semantic interpretation in a particular domain. The methodology generalizes the techniques used in existing programs for molecular biology, oncological pharmacology, and anatomy.

Indexing Initiative

The Indexing Initiative project investigates methods whereby automated indexing may partially or completely substitute for expert indexing of the biomedical literature by humans. The project is pursuing concept-based indexing methods that go beyond automatic word-based indexing and will be considered a success if its retrieval performance is equal to or better than that of systems using humanly-assigned index terms. During the past year team members tested a prototype indexing system based on three fundamental indexing methodologies. The first of these calls on the MetaMap program to map citation text to concepts in the UMLS Metathesaurus. The second approach, the trigram phrase algorithm, uses character trigrams to match text to Metathesaurus concepts, while the third uses a variant of the PubMed related citations algorithm to find MeSH headings related to input text. Results from the three methods are restricted to MeSH and combined into a ranked list of recommended indexing terms.

Retrieval experiments to determine the adequacy of the Indexing Initiative's system will be performed in the coming year. In addition,

plans for applying its results to both semi-automatic and fully automatic indexing environments are being developed. Research into the system's indexing methods will continue. In particular, a major word sense disambiguation effort based on journal descriptor indexing is being undertaken to resolve ambiguities encountered during the automatic indexing process. Finally, the Indexing Initiative team will be extending its research to address the full text documents that are becoming increasingly available.

Terminology Server

There is often a mismatch between the vocabularies of users and the vocabularies of clinical and information retrieval systems. The purpose of the Terminology Server project is to provide tools to bridge that gap. For example, health care consumers can not be expected to know the technical vocabulary of medicine. Functions of the terminology server should help convert terms provided by a user into an appropriate biomedical terminology that can be utilized by a system. The terminology server is a set of middle-ware components that should enhance user-to-system communication by providing, for example, synonyms, lexical variation including spelling information, and more generally, the knowledge associated with terms. Thus, it bridges the gap between levels of abstraction and between dialects and disciplines. It can also be used to translate between vocabularies or specialized languages.

Ongoing research includes evaluating the terminology server functions that were implemented for the ClinicalTrials.gov system and studying new methods to support improved information retrieval through terminology servers. Currently, we are reviewing methods to provide developers with options for selectively filtering the UMLS Metathesaurus, based on their needs. These options would provide more flexibility in limiting the types of terms (e.g., preferred terms only) and source vocabularies used in a terminology server. Finally, we are developing stand-alone terminology server modules that can be incorporated into any medically-oriented application.

Just-In-Time Information

Although the discoveries made in the biomedical research programs of the world are impressive, the frequency with which these discoveries are incorporated into routine clinical practice is disappointingly low. It is a paradoxical irony that the magnitude of the advances produced by the biomedical research community has overwhelmed clinicians with more information than they can absorb effectively. As a result of this information overload, clinicians find it difficult to answer the questions that occur in the routine care of patients, and cannot offer the most advanced treatments to their patients.

The Just-In-Time (JIT) project is an attempt to build upon NLM's biomedical information databases and construct a real-time; Internet based information system that provides succinct, highly relevant information to clinicians at the point of care. It will incorporate the literature found in MEDLINE with NLM databases that contain clinical guidelines and ongoing clinical trials. The components of the JIT research agenda include study of the structure of physician questions, improving database search strategies, and developing appropriate ranking hierarchies for medical information.

The JIT project is currently in the process of modeling questions of clinicians. In a collaborative process with several academic medical centers, a database of clinician questions is being designed that will be used to study the content and structure of clinician questions and to construct "generic queries." These generic queries will then be used as templates, into which clinicians will insert the specific topic of their question. Each template will also be linked to a unique preformatted ("canned") search strategy or "hedge," which will allow multiple NLM databases to be searched and will increase the sensitivity of the searches. This strategy of combining a preformatted generic question template with a predetermined "hedge" with a specific topic, supplied by a clinician, ensures that well structured questions are used to initiate a search, eliminating one of the common reasons that database searches are unsuccessful.

Proteus Project

An investigation was started in the design of system architecture for using medical knowledge in the form of executable distributed components to construct clinical protocols and thereby to represent the clinical process. The goal is to develop a system that supports medical decision making, data entry and data storage in a clinical setting. In this approach, called Proteus (PROTOCOLS Editable by USERS), clinical processes are represented by three types of "knowledge components": actions, processes and events. Each component has a mechanism to infer its own value and to infer which next action has to be launched.

A Java-based proof-of-concept module (Protean) based on the Proteus architecture has been built. As a way to demonstrate its operation, a clinical protocol Magnesium Sulfate therapy for severe pre-Eclampsia/Eclampsia which can be run in Protean has been created. Protean was demonstrated at the Medical Informatics Trainees meeting at NLM in July. In September we presented a description of the architecture and demonstrated Protean at the SmartSystems 2000 conference at Houston, TX sponsored by NASA, University of Texas-Houston, and the National Space Biomedical Research Institute.

Smart Cards

NLM sponsored or cosponsored several projects involving smart card technology during the past year. A smart card is a credit card sized plastic card with an embedded circuit chip. The chip can be a microprocessor with internal memory capable of running small programs, or simply a non-programmable memory chip.

An NLM-sponsored project at the Concurrent Engineering Research Center (CERC) at West Virginia University uses smart cards for authentication and data storage in rural healthcare delivery. Both patient cards and health provider cards are used. NLM has been a cosponsor of the Western Governors' Association Health Passport Project. This project involves the storage of data from multiple Federal, state, and local agencies on cards used by clients receiving benefits such as

well child care, checkups, immunizations and food benefits.

A Medical Informatics Fellow recently explored the possibility of storing on an advanced Cyberflex Java Card information specific to cardiac emergencies for patients known to be at high risk. If successful, this will provide a means of carrying particularly important information in a portable, updatable record for a specific patient population.

Information Systems

Consumer Health Informatics

ClinicalTrials.gov is a new consumer health informatics application that was developed by the Center in response to legislation requiring the NIH to create a database of clinical trials information. Increasingly, people are turning to the Internet to look for answers to their health questions. This raises a number of research questions, including the type of content that should be created and how that content can be put into the appropriate medical context. The structure of the ClinicalTrials.gov application was designed to accommodate these issues.

ClinicalTrials.gov provides patients, families, and members of the public easy web-based access to current information about clinical research studies. Each record in the database includes summaries of the purpose of the clinical research study, together with the recruiting status, the criteria for patient participation in the trial, the location of the trial, and specific contact information. Other information that may help a patient decide whether to enroll in a particular trial includes the research study design, the phase of the trial, the disease or condition, and the particular drug or therapy under study. An important feature of ClinicalTrials.gov is that it provides links to other online health resources such as MEDLINEplus that help place clinical trials in the context of a patient's overall medical care.

Nine months after the release of ClinicalTrials.gov in February 2000, the site had received over 15 million hits, with an average of over 60,000 daily connection requests from over 3,000 individual computers each month. The site

has generated a great deal of press interest and positive feedback from users. ClinicalTrials.gov contains over 5,000 trials sponsored primarily by NIH Institutes. In the coming year, the database will be expanded to include trials sponsored by other Federal agencies as well as the pharmaceutical industry. Work continues on an updated data entry tool to facilitate submission of trials to the site.

Digital Library Research

Digital library research involves all aspects of creating and disseminating digital collections, including standards, emerging technologies and formats, copyright and legal issues, effects on previously established processes, protection of original materials, and permanent archiving of digital surrogates. Research issues currently in focus are long-term preservation of digital archives, innovative methods for creating and accessing digital library collections, the development of modular and open information environments, interoperability among digital library systems, investigation of the role of well-structured metadata, and the exploration of different "points of view" on the same underlying data set.

In the fall of 1998 the Profiles in Science web site was released. The site uses innovative digital technology to make available the manuscript collections of prominent biomedical scientists of the twentieth century. The content of Profiles in Science is created in collaboration with NLM's History of Medicine Division, which processes and stores the physical collections. The materials have been donated to NLM and contain published and unpublished materials, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audio tapes and other audiovisual materials. Presently Profiles in Science features the collections of five prominent American biomolecular researchers: Oswald Avery, Joshua Lederberg, Martin Rodbell, Julius Axelrod, and Christian Anfinsen.

This year the Digital Library Research team also successfully migrated an early digital library, the Regional Medical Programs (RMP) collection, into a more modern digital library

consistent with Profiles in Science. The heart of the RMP collection is approximately 40,000 pages comprising some 1,500 documents related to the Regional Medical Programs of the 1960's and 1970's.

NLM Gateway

The National Library of Medicine offers an increasing number of Internet-based information resources, each with its own user interface. Lister Hill Center researchers have created the NLM Gateway to let users initiate searches in multiple retrieval systems from a single interface. The target audience for the new system is the Internet user who comes to NLM not knowing exactly what is available or how best to search for it. The NLM Gateway entered beta testing in early summer and was released in October 2000.

The initial version of the NLM Gateway offers access to the following online resources:

- MEDLINE (includes PREMEDLINE)
- OLDMEDLINE
- LOCATORplus online catalog information for books, serial titles, audiovisuals
- AIDS Meeting abstracts
- HSRPROJ health services research projects
- MEDLINEplus consumer health topics information
- MEDLINEplus consumer drug information
- Document delivery through the Loansome Doc system
- UMLS Metathesaurus.

Gateway users enter a query once. The query is reformulated and sent automatically to multiple retrieval systems having different characteristics but potentially useful results. Results from the target systems are presented in categories (for instance, journal article citations; books, serials and audiovisuals; conference abstracts; databanks; consumer health information) rather than by database. In some categories, multiple collections are searched.

Online visitors are invited to use the Gateway for an overview scan of some of

NLM's resources. Some users will find what they need immediately. Others may find that one resource such as PubMed or MEDLINEplus has information they'd like to know more about. They may then choose to go straight to that resource for a focused search using its native interface. Direct links to other major NLM resources are provided from the Gateway's search screen. This combination of a single point of access for an overview scan coupled with focused searches available for a second phase of inquiry should help improve user access to information offered at NLM's expanding series of Web sites. The Gateway will replace the earlier Internet Grateful Med system.

HSTAT

Information for HSTAT (Health Services/Technology Assessment Text) is received from multiple government agencies. The most heavily used collections are the Consensus Program and the Clinical Guidelines. A DataTool program for facilitating the process of validating the SGML-encoded documents received and adding them to the system is being beta tested. A total of 54 new documents have been submitted for the HSTAT collection. The Clinical Center Protocols collection was removed from HSTAT; it is now covered by the ClinicalTrials.gov system. A subject list has been created so that HSTAT documents can be displayed under subject headings as well as grouped by supporting agency.

Several new functions have been added to the HSTAT full-text retrieval system. A means of using software agents for expanding queries with synonyms has been completed and is undergoing testing. Agents for results processing including ranking and for spelling correction are under development. Integration of the several agents into an agent system allowing them to communicate with each other has been completed. The agent technology will be incorporated into areas of the NLM Gateway when appropriate. The new HQuest client for HSTAT written in Java and using servlets is essentially complete. A new version of HSTAT based on the Versant object oriented database management system and accessed by the HQuest

client will be tested during the first quarter of 2001.

Medical Education and Outreach Activities

Phase I of the "Breath of Life" DVD was completed on schedule and had a public showing at the annual meeting of the American Association of Physicians' Assistants. The conference, which ran May 27 to June 1, 2000, was held at the McCormick Place Convention Center in Chicago, Illinois. The highly interactive program incorporates digital video, audio, 2D graphics, and advanced three-dimensional graphics and animations to show the entire scope and content of the Breath of Life exhibition. One hundred and twenty minutes of full-screen, full-motion video can be navigated from the single DVD disc. Program design allowed for added enhancements to several aspects of the current exhibition. Additional on-camera interviews and video stock materials enabled the Faces of Asthma section to feature more fully developed, compelling individual stories of people who are successfully managing their asthma. The program also features a 6-minute video overview, offering a comprehensive and thematic orientation to the entire exhibit. As the DVD standard allows for 5.1 Dolby digital encoding, an original introductory animation was developed with accompanying audio produced for enhanced surround sound delivery.

The Movement Disorders Video Database Project was a collaborative project with Yale University School of Medicine's Movement Disorders and Neurodegenerative Diseases Clinic, the Center for Advanced Instructional Media and the Biomedical Communications Department. This pilot effort established a digital video database of high quality, full-motion video of medical significance. Neurologically based movement disorders were selected as subject matter which would best be characterized by video and audio.

Project collaborators at the Yale University School of Medicine launched the initial set of five patient data segments on November 10, 1999 at a symposium for nurse practitioners and physicians' assistants, held in New Haven, Conn. Lister Hill Center staff

provided the five edited, compressed video sequences, and Dr. Carl Jaffe of Yale developed the HTML interface and presentation front-end. The database was presented during two conference plenary sessions, "Parkinson's Disease: Diagnostic Challenges and Management Controversies," and "Treatment and Motor Fluctuations." Participant feedback was very positive. In particular, the health professionals commented on the significant advantage of this interactive tool over conventional video.

The full video data set of Parkinson's patients was completed and delivered to collaborators at the Yale University School of Medicine. This included 15 patients with varying degrees of Parkinson's disease, in 17 sequences, including 3 pre-medication and post-medication sequences. In developing the first complete set of Parkinson's patient video data, Center staff worked to develop optimum video and audio compression schemes for cross-platform compatibility, and enhanced functionality and usability by incorporating chapter titles and tracks.

Several videotapes of NLM-supported telemedicine projects were produced this past year. In December, location videotaping of the University of Missouri telemedicine project was completed. A videoconference between a physician at University Hospital in Columbia and staff and a patient at the Loch Haven Nursing Home (60 miles away) was recorded. Interviews were also conducted with key members of the University of Telemedicine project, the nursing home administrator, and a local Macon physician.

A 48-minute videotape "Health Services Research—A Historical Perspective" was completed for showing at the Annual Meeting of the Academy of Health Services Research and Development in Los Angeles. This project required extensive pre-production, production, and editing effort including scripting, many interviews, historical picture research, and source materials retrieval from Presidential Libraries, the National Archives, the Library of Congress and Johns Hopkins University.

The National Museum of Health and Medicine and George Mason University are working to create a three-dimensional atlas of

the human embryo based on the historic Carnegie Collection of the Human Embryo at the Museum. Milestone events of the project are to be documented on videotape. The initial taping in April was of a peer review meeting at the museum. During the meeting, experts from several institutions around the country made selections from slides in the collection. Two milestone segments were also videotaped in August including the actual automated scanning of the slides in the collection. Some slides being used in the project date back to the beginning of the 20th century. Additionally, the "Site Visit" featuring reports from the project participants was videotaped. The meeting took place at the Lister Hill Center with offsite participants in San Diego and Seattle participating via video teleconferencing.

The Learning Center for Interactive Technology was redesigned and transformed into the Collaboratory. The lab investigates innovative means for assisting health science institutions in their use of online distance learning technologies. A database of Internet accessible health professions education materials, EtherMed, was developed. It is intended for use as a research tool and a mechanism for sharing online courseware among health professions schools. Additional references and links are being added by colleagues at the University of Utah and the University of Alabama at Birmingham.

Experiments with the Lucent video conferencing technology continued as well as a new set of experiments using Litton conferencing technology. Collaborators include NASA, the University of Alabama at Birmingham and Trinity University in Dublin, Ireland. Interactive demonstrations were presented over the Abilene Network from the Collaboratory to the I2 Conference in Washington, DC and to the Slice of Life Conference in Salt Lake City, Utah.

Image Processing

Visible Human Project

The Visible Human Project data sets are designed to serve as a common reference point 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 being made available through a free license agreement with the NLM. They are being distributed to licensees over the Internet at no cost, and on DAT tape for a duplication fee. The data sets are being applied to a wide range of educational, diagnostic, treatment planning, virtual reality, artistic, mathematical and industrial uses by over 1400 licensees in 42 countries. The Visible Human Project has been featured in more than 800 newspaper articles, news and science magazines, and radio and TV programs worldwide.

The data sets are having their greatest effect on health care and health education and thus benefit the general public. The data sets are used as a normal reference and as an aid in the diagnostic process. Programs under development will be used to educate patients about the need for and purpose of surgery and other medical procedures as well as to permit physicians to plan surgery and radiation therapy. The images from the Visible Human data sets are used in several prototype virtual reality surgical simulators. Educational materials that make use of the Visible Human data sets are beginning to be used by students from kindergarten to practicing health care professionals. The data sets are being used to form the basis of interactive games to entertain as well as to educate. Automobile manufacturers now include passenger injury models based on Visible Human data to their vehicle crash simulation models. Engineers and physicists are creating models to quantify human exposures to various forms of electromagnetic radiation. The data are also being used by mathematicians as an application for what were previously only theoretical mapping theories. Several artists are using the data set as the basis for new multi-media art forms.

Online demand for the Visible Human data has remained high since its availability. Both the number of users accessing the FTP site and the number of files retrieved have shown a continued interest in the data set. The data set contains 70 Gbyte of uncompressed full color and radiological images. Image files are stored

in a compressed format in directory structures for the male and female images. Each of the main directories divides into subdirectories for the full color, MRI, and CT images. The larger digital color subdirectory is further divided into anatomical regions, head, thorax, etc.

FTP access to the data is over a T3 Internet distribution node which is internally connected to a 100 Mbs local area network, with the data stored as Unix files on a Sun SPARCServer. The data set is stored on a SPARCstorage Array in a stripped data block configuration, with mirrored volume copies.

The AnatLine anatomical digital image database was released for beta testing beginning April 2000, and it continues to be accessed by an international group of biomedical users. AnatLine is the online delivery database component of the 3DSystems anatomical image database management system. In addition to its online delivery capability, the image management system also includes a larger catalog database which maintains the archival data records used for processing voxel structures, generating 3D rendered images, and maintaining anatomical and spatial relationships among its image data and file structures.

The core capability of the system is its anatomical image database which provides a query and retrieval engine to access high resolution image records. Four types of image records are available; anatomical cross sections with labels, volume of interest, segmented masks, and rendered images. Presently, the user database stores images and structures for the thorax of the Visible Human male. These anatomical structures of the thorax were extracted from the Visible Human male 70mm color film data set. Visible Human anatomical images prior to the development of this database were only made available over the Internet via FTP as raw files containing gross anatomy cross-sections of the human body, requiring complex processing to extract individual anatomical structures. AnatLine extends this capability by providing users the ability to query and retrieve selected anatomical structures

The first phase of the evaluation of the high resolution scanning of the Visible Human Female 70mm film images was completed and the final test files and report were delivered.

After evaluating the test images scanned at resolutions of 3,300 to 12,200 ppi, it was decided that a resolution of 4,450 ppi was sufficient to capture the available information within the VHF transparencies. A total of 126 images were then scanned at the 4450 ppi and made available for review. The methods used for scanning proved to be acceptable and will be employed for the entire film.

The second phase of the project, to scan all of the Visible Female 70mm film images, is now under way. The complete set is being scanned by JJT Consulting in Austin, TX. They will digitize 5189 film images at 4500 ppi and 16 bits per color channel. The resulting file size of these images will be approximately 450 Mbytes. JJT Consulting has developed its own software for the Windows NT platform to quickly open and display these large files. Multiple derivative images will also be provided at lower resolutions.

The Image Storage and Transmission Optimization project involves research into compression and transmission techniques to improve access to, and delivery of, data-intensive biomedical images, with specific focus on the Visible Human color image set. The CCD captured male and female data in the Visible Human amounts to 55 GB, but the 70mm photographs taken during cryosectioning the cadavers, currently being scanned at much higher resolution, will yield a total of about 235 GB. Since datasets of this size will strain both storage and transmission resources, this project is undertaken as an investigation of both compression and advanced communications techniques to alleviate these problems.

This year the main focus has been in image compression, and we have made progress in creating mechanisms for the access, retrieval and display of compressed images. The compression research has investigated both lossy and lossless techniques, since both classes have applicability: lossless techniques would serve the objective of retaining all the original image data in storage; however, since much higher compression ratios are possible with lossy techniques, these would serve our objectives of rapid transmission over the Internet.

Studies conducted in lossy compression techniques included wavelet transformation for

image decomposition and both scalar and vector quantization for data reduction. Starting with the well established premise that the fidelity of a reconstructed image can be predicted once the image statistics are known, the probability distribution functions (pdf) of a number of wavelets and selected filter lengths providing the best fit to a generalized Gaussian distribution for the images were investigated. The filter length is a crucial factor in shaping the pdf and hence in predicting achievable bit rate at minimum distortion in a quantization scheme. The pdfs of the coefficients in wavelet transformed subimages of selected Visible Human slices were investigated using filter taps from 4 to 20 to establish the optimality of specific wavelets for this class of images, and it was concluded that Daubechies and Symlet wavelets with 12 filter taps introduced the least variance from the generalized Gaussian distribution function for optimal coding at all levels of decomposition. Using this filter length (DAUB 12), a detailed analysis was undertaken of the compression performance possible with a variety of techniques.

Progress was made toward developing the tools for accessing, retrieving and viewing the compressed Visible Human images. First, a Java applet, Image Retriever, was developed for selecting one or a range of slices clicking on a simplified body map created from the data set, establishing the TCP/IP connection between client and server, and saving the compressed image in the user computer memory. Second, a Java application, Image Viewer, was created to allow the user to choose the image(s) to decompress, run the C code for inverse EZW (Embedded Zerotree Wavelet) and AAC (adaptive arithmetic coding) transformation, and display the image in its raw form (the form it exists in the original data set) without requiring it to be in GIF or JPEG form

MARS: Automating the Production of MEDLINE Records

This project (Medical Article Record System, or MARS) aims to develop automated systems for the extraction of citation and abstract data from medical journal articles to create bibliographic records for the MEDLINE

database. The design of these systems is founded on research in document image analysis and understanding, database design, artificial intelligence using rule-based and artificial neural network approaches, graphical user interface design, speech recognition, image processing, and related areas. The production of MEDLINE records has traditionally been done by manual keyboarding. The first system, MARS-1, combined the keyboarding of citation data (journal name, date, author, title, affiliation, page numbers, etc.) with scanning and automatic text conversion by optical character recognition (OCR) of the abstracts which, if keyboarded, proves very labor-intensive.

While gradually improving and maintaining MARS-1 in routine production, research staff conducted R&D toward more comprehensive automation, first focusing on the design of a database-centered and database-driven next generation MARS-2 architecture with the definition of over 130 tasks. The initial tasks were concerned with the SQL Server database design, GUI issues and algorithms for automated processing. These included: the development of an entity relationship diagram, a data dictionary, and C++ database read/write classes; a daemon to interface the database to the OCR system; a scanner interface to the database and a new scanner GUI design incorporating a magnifying glass feature, a dialog box for confirming MRI (journal issue identifier), and a continue button for abstracts that span two pages; and comprehensive reconcile workstation software. Also, an initial top-level design for the recognition of Greek and other biomedical symbols is completed and work has begun on the design of a subsystem that will implement a feature extraction and matching classifier for these "special" symbols.

The second generation system, MARS-2, was developed and placed into operation, first at NLM, and in June 2000 moved to an offsite facility in downtown Bethesda. Besides the automated subsystems (daemons), MARS-2 has four types of operator workstations: scanner, Edit, Reconcile and Admin. Edit is for entering information that is not extracted automatically; Reconcile is to verify the accuracy of the whole record prior to uploading it to the NLM database that retains

records for subsequent indexing; and Admin for the operations supervisor to monitor the quality of the work and return the data created to an earlier stage for reworking.

While both MARS-1 and MARS-2 systems are currently operating in parallel at the new site, the goal is to eliminate MARS-1 eventually. Research toward more comprehensive automation focuses on the development of algorithms and software modules for automated zoning, automatic field identification (labeling), intelligent spellcheck, and for reformatting field syntax.

DocView Project

The goal of the DocView project is to apply document image processing R&D to document delivery via the Internet. Furthermore, the project addresses NLM's mission of providing document delivery to end users and libraries, and incorporates advanced digital imaging techniques. The first product of this research is a widely distributed client software that enables end users to directly receive documents from Ariel systems or web sites. While Ariel, a product of the Research Libraries Group, is used by libraries and document suppliers routinely to send documents via Internet to similar organizations, there are few options for end users to directly receive them. The DocView client software fills that niche. In its 31st month since release, DocView has been downloaded by 9,254 registered users in 102 countries, the latest being Romania and Bulgaria.

The DocView client software, which runs under any version of Microsoft Windows, enables an end user to receive documents over the Internet at the desktop, retain them in electronic form, view the images, organize the received documents into "folders" and "file cabinets," electronically bookmark selected pages, manipulate the images (zoom, pan, scroll), copy and paste images, and print them if desired. DocView also serves as a TIFF viewer for compressed images received through the Internet by other means, such as Web browsers. Users may receive document images either via Ariel FTP or Multipurpose Internet Mail Extensions (MIME) protocols. Using DocView,

users may also forward documents to colleagues for collaborative work. Major users appear to be scientists receiving documents via the Internet at their desktops from Ariel workstations at their libraries. This was the case, for example, for researchers at a malaria research site in Kilifi, Kenya who used DocView to receive medical articles from an Ariel system at NLM.

A document conversion server (DocMorph) is currently in beta-test. DocMorph allows users to upload image files for conversion to alternative formats, a recommendation by many DocView users who often find that the image files they receive cannot be used when they do not have suitable viewer software. It lets users and librarians to convert TIFF images to PDF or other formats that may be preferable in their application. The DocMorph Server has also been speech-enabled ("Reading Room" function), allowing a user to upload a document image and have it converted to speech so that, through OCR and speech synthesis, the document can be read out to the user.

Biomedical Imaging and Multimedia Database R&D

The goal of this program is to address fundamental questions that arise in the handling, organization, storage, access and transmission of very large electronic files in general and digitized x-rays in particular. A special focus is research into these topics as applied to heterogeneous multimedia databases consisting of both images and text. This work has evolved from a previous project named DXPNET, conducted in collaboration with two other agencies, the National Center for Health Statistics and the National Institute of Arthritis, Musculoskeletal and Skin Diseases.

WebMIRS is a Java applet that allows remote users to access data from two surveys conducted by the National Center for Health Statistics. These are the National Health and Nutrition Examination Surveys II and III (NHANES II and III), carried out during the years 1976–1980 and 1988–1994, respectively. The NHANES II database accessible through WebMIRS contains records for about 20,000 individuals, with about 2,000 fields per record;

the NHANES III database contains records for about 30,000 individuals, with more than 3,000 fields per record. In addition, the 17,000 x-ray images collected in NHANES II may also be accessed with WebMIRS and displayed in low-resolution form. WebMIRS allows a user to control a graphical user interface to construct a query of the NHANES II or NHANES III data. Beta testing began this year and is ongoing, with testers not only in the United States, but also in Korea, Sweden, and Mexico. WebMIRS was used in two semesters of a graduate course in public health statistics at Columbia University in 1999–2000 to demonstrate new technological data access methods, and a real time data acquisition and analysis was demonstrated using WebMIRS at the CDC Data Users Conference in Bethesda, Md. in July 2000.

The Digital Atlas of the Spine is a dataset of cervical spine and lumbar spine images with interpretations validated by a consensus of medical experts, along with software to display and manipulate the images. The images in the Atlas were chosen from the 17,000 images collected in the NHANES II survey. We convened two workshops in collaboration with other NIH researchers to seek expert advice and consensus on a wide set of technical and biomedical issues related to the radiological interpretation of this set of images. Among the issues covered were the exact features to be interpreted. Radiographic features considered for interpretation of the cervical images were anterior osteophytes, posterior osteophytes, disc space narrowing, sclerosis, vacuum phenomenon, and spondylolisthesis. For the lumbar images, features considered included anterior osteophytes, posterior osteophytes, disc space narrowing, sclerosis, vacuum phenomenon, spondylolisthesis, spondylolysis, and DISH. A subset of these features was selected as likely to be consistently interpretable from the NHANES images. This selection of features, based on the consensus of experts at the workshop, took into account published studies relating to the likelihood of obtaining consistent readings for the features considered. The features identified by the workshop as consistently readable were those chosen for the Atlas.

Medical Informatics Training

The Medical Informatics Training Program provides training for students at various stages of their careers and brings talented people to the Center. The program recruits promising students into careers in medical informatics, playing a role in developing researchers and leaders for the field. Potential areas of research for participants include digital library research, automated indexing techniques, vocabulary and thesaurus research, medical language processing, image processing, document analysis, belief networks, wide area network technologies, client/server design, database design, machine learning, expert systems, and computer-based learning.

This past year, the Lister Hill Center provided training to 45 participants from 14 states and 9 countries. The participants included 4 high school students and teachers, 17 undergraduate students, 12 graduate or medical students, 6 postdoctoral or post-MD fellows, and 6 visiting faculty scholars. Students worked closely with NLM staff and presented the results of their work to the NLM community.

The NIH Clinical Elective in Medical Informatics, one of more than 20 rotations at the NIH, was again held at NLM in March and April 2000. The Center trained five third-year and fourth-year medical students from U.S. and British medical schools. The elective provided an overview of medical informatics through outstanding lectures both from invited speakers and from NLM staff and through two field trips to area medical facilities. Students had the opportunity to work closely with NLM and NIH preceptors on independent research projects, presenting their work in a series of seminars at the end of the elective.

The Center participates actively in programs supporting minority students, including the Hispanic Association of Colleges and Universities (HACU) and the National Association for Equal Opportunity in Higher Education (NAFEO) summer Internship programs.

Office of the Public Health Service Historian

The staff of the Office of the Public Health Service Historian delivered a number of lectures and published several papers on research interests during the year. The PHS Historian has been working with the Office of the Surgeon General (OSG) to develop a more accurate list of Surgeon General's Reports since 1964, to compile a set of the Reports for OSG, and to explore with the NLM the possibility of scanning the reports and making them available on the *Profiles in Science* system. The Historian has also been working with the Office of Public Health and Science (OPHS) on an article dealing with the history of Surgeon General's Reports. The Historian and historical consultant Dr. Caroline Hannaway conducted an oral history interview with Dr. Eric Goosby, Director of the Office on HIV/AIDS Policy, OPHS, as part of an ongoing project to document the history of AIDS policy in the PHS. The office also conducted an oral history interview with John Eason, Jr., the first African American to be commissioned in the PHS. The Office developed a small exhibit on the Surgeon General's priorities for display at the Reserve Officers' Association. The Historian and his staff have also been cooperating with the NIH History Office on their project to build a database of past NIH employees. A significant amount of time was spent in planning and implementing the 2000 meeting of the American Association for the History of Medicine. The Office provided information to OPHS for use in speeches and writings of the Surgeon General, Dr. David Satcher.

Resource Support and Development

Audiovisual Support

Center staff support the audiovisual requirements of NLM's educational and information programs. With the mission requirement of the NLM expanded to include effective outreach activities, the range of support

that the Center provides to these programs continues to increase. From the application of optical media technologies and teleconferencing to support for Web design, the graphics, video, and audio materials requirement has increased in quantity and diversified in format. Staff investigate image quality and resolution, color fidelity, media transportability, media storage, and visual communication. The facilities and hardware systems must reflect state-of-the-art standards in a rapidly changing field. High definition video is a development area being explored that represents the future for improved electronic image quality. Multimedia systems and techniques, visualization and networked media are being pursued for the educational and the cost advantages that they offer. Three dimensional computer graphics, animation techniques, and photorealistic rendering methods have changed the tools and products of the graphic artists in the Center. Digital video and image compression techniques are central to projects being pursued in areas of image storage and transmission.

Emerging Network Retrieval Protocols

This project continued to develop a collection of software tools related to the use of Unicode, including a Java-based graphical user interface to the multilingual International Classification of Primary Care (ICPC) vocabulary from WONCA, the World Organization of Family Doctors. ICPC was recently incorporated into the UMLS Metathesaurus. The project intends to develop a Web-based environment for open collaborative maintenance and enhancement of this 20-language vocabulary. This will be based on the updated Apache/MySQL/PHP Web server application environment put in place by the group over the past year.

The group maintains several online Web services, including multimedia public exhibits and tools for internal use by Center personnel. The "depot" shared local software repository was expanded with numerous tools related to UNIX and network security.

System Security and Advanced Network Planning

This group's work during the year concentrated on computer security, the Lister Hill Center network, and the Next Generation Internet. Computer security involved refinements of access controls and the development of a security classification organization. A secure-subnets working group is developing a classification of NLM systems that would be used to define and implement different levels of network access between NLM and the Internet.

Work on the Lister Hill Center network has continued with the development of a Gigabit backbone. The current Fast Ethernet network is based on a star configuration of Cisco 5500 switches, each connected to the Computer Science Branch 5500 switch. This configuration is being replaced by one centered on a Foundry Gigabit switch connected to the 5500 switches by Gigabit links. The Foundry switch will also support Gigabit connection from some of the larger servers.

NLM connected to the vBNS (very high speed Backbone Network Services), an NGI network, on January 9, 2000. In the third quarter of 2000, we connected to the Abilene network and more recently to the NGIX-DC. While the vBNS was the first NGI network to connect universities and other research institutions, many of the vBNS constituents have moved to the Abilene network. The third connection, NGIX-DC, is a connection point for the Federal NGI networks.

The vBNS connection provided the opportunity to compare the performance of some typical applications over the Commodity Internet and the vBNS. A preliminary study showed that normal web applications performed only slightly better over the vBNS than over the Commodity Internet. These web applications involved small amounts of data for each transaction. Other applications involving large amounts of data per transaction performed significantly better over the vBNS. This study pointed to avenues for further research in application development and performance measurements.

Next Generation Internet: Infrastructure Development and Applications

Through the technical and administrative efforts of Lister Hill Center engineering staff in 1999 NLM entered the Next Generation Internet with OC-3 (155 Mbps) linkage to the high speed vBNS network, thereby connecting to over 88 major educational and research institutions on vBNS, in addition to peering with other high speed networks allowing connections to a number of other national and international centers. The backbone speed of vBNS increased from its 1999 figure of 622 Mbps to the current rate of 2.5 Gbps. In 2000, we moved to vBNS+ to allow more flexibility of operation by avoiding the authorized use policy limiting our use of vBNS. In July 2000, NIH officially became a participant in the Abilene network, the backbone network for Internet2 (a university-centered initiative.) Connectivity to Abilene is through the network services of the Mid-Atlantic Crossroads (MAX) gigapop located at College Park, MD. In connection with this infrastructure development, Center staff represent NLM at the Joint Engineering Team (JET), a coordinating body overseeing large-scale Federal networking activities. The JET coordinates operations and engineering planning among several Federal agency networks and the Internet2, and deals with issues such as network access points, international connections, traffic monitoring, performance measurement, multicast distribution, and the deployment of IPv6 and Quality of Service (QoS) capabilities.

Maryland Governor's Task Force on High Speed Networks

In 2000, the Lister Hill Center continued to serve as a federal representative to the Maryland Governor's Task Force on High Speed Networks and the Engineering Advisory Group. The Task Force developed a comprehensive plan for bringing the state's network infrastructure in line with the needs of the 21st century. This plan, completed and presented to the legislature, contains recommendations to combine existing state resources to maximize the state's return on investment; use existing state owned fiber where available; use current right-of-ways the state

possesses to add additional fiber in underserved regions such as the Eastern Shore, Western and Southern Maryland; provide equity of access to all regions of the state, and support multiple segments of our society; promote collaboration among businesses, educational institutions, governmental bodies and research institutions; and conduct a select number of high priority pilot projects in health care, business infrastructure development, and state government functions. A major contribution by the Lister Hill Center was made in the development of pilot projects in health care involving remote oncology treatment planning and remote intensive care support.

Engineering Laboratories

Document Imaging Laboratory

This laboratory supports DocView, MARS and other research and design projects involving document imaging. Housed in this laboratory are advanced systems to electro-optically capture the digital images of documents, and subsystems to perform image enhancement, segmentation, compression, OCR and storage on high density magnetic and optical disk media. The laboratory also includes high-end Pentium-class workstations running under Windows 98 and NT, all connected by 100 Mb/s Ethernet, for performing document image processing. Both inhouse developed and commercial systems are integrated and configured to serve as laboratory testbeds to support research into automated document delivery, document archiving, and techniques for image enhancement, manipulation, portrait vs. landscape mode detection, skew detection, segmentation, compression for high density storage and high speed transmission, omnifont text recognition, and related areas.

MARS Production Facility

This off-campus facility houses high-end Pentium workstations and servers that constitute MARS-1 and MARS-2 production systems used by operators to scan medical journals and also keyboard data to produce MEDLINE records. While primarily for

production, this facility also serves as a laboratory to collect data for the continual improvement of the MARS systems, such as a large collection of bitmapped document images, zoned images, labeled zones, and corresponding OCR output data. This collection serves as a test set for research into techniques for autozoning, autolabeling, autoreformatting, intelligent spellcheck and other key elements of MARS.

Image Processing Laboratory

This laboratory is equipped with a variety of high end servers, workstations and storage devices connected by 100 Mb/s Ethernet. The laboratory supports the investigation of image processing techniques for both grayscale and color biomedical imagery at high resolution. The laboratory has computer and communications resources and image processing equipment to capture, process, transmit and display such high-resolution digital images. Most machines are equipped with multiple networking ports (FDDI, ATM, Ethernet, fast Ethernet) which allow, in addition to standard networking capabilities on the local Ethernet, the capability of alternate physical communications channels with these machines. This capability has been used in communications engineering experiments for point-to-point satellite channels connecting these machines with remote sites. ATM switches connect the Ethernet and FDDI networks to other local area networks throughout the building, to the Internet, and to experimental ATM networks such as ATDnet and MCI's research network, in addition to vBNS, the infrastructure for the Next Generation Internet and Internet-2 initiatives.

External Research Support

Telemedicine

The Telemedicine Program is designed to evaluate the impact of advanced networking on health care, research, and public health; to test methods to preserve the privacy of individual health data while also providing efficient access for legitimate health care, research, and public health purposes; and to assess the utility of emerging health data

standards in health applications of advanced communications and computing technologies.

NLM is the lead agency within DHHS for the government's High Performance Computing and Communications initiative and as such has a direct interest in the use and effects of advanced networking on health care. The growth of the Internet and the increasing access to high-speed computers and communications by consumers, health care providers, public health professionals, and basic, clinical, and health services researchers is having a fundamental effect on health and human services throughout the nation. Major research and evaluation issues included in NLM's telemedicine program arising from the current and future impact of advanced networking include the impact of telemedicine on the health care system as a whole and on cost, quality, and access to care for specific populations; the benefits of integrated access to practice guidelines, expert systems, bibliographic databases, electronic publications, and other knowledge-based information from within computer-based patient record systems and other automated systems that support research and practice; the maintenance of patient confidentiality as increasing amounts of electronic health data are transmitted via telecommunications during health care and aggregated for important public health and research purposes; and the development of data standards and uniform practices for effective transmission, aggregation, and integration of health care, public health, and research data.

Over a five-year period, 19 telemedicine projects, affecting rural, inner-city, and suburban areas, with a total budget of \$42 million were funded. The projects, located in 13 states and the District of Columbia, were designed to serve as models for evaluating the impact of telemedicine on cost, quality, and access to health care; assessing various approaches to ensuring the confidentiality of health data transmitted via electronic networks; and testing emerging health data standards. Each project was required to review and apply the recommendations from two NLM sponsored National Academy of Sciences (NAS) studies: one on criteria for the evaluation of telemedicine (*Telemedicine: A Guide to Assessing Telecommunications for Health Care*); and the other on best practices for

ensuring the confidentiality of electronic health data (*For the Record: Protecting Electronic Health Information*).

During this past year most of the contracts under the telemedicine program came to an end. A symposium is planned for March 2001 at which each of the participants will make presentations about lessons learned.

Visible Human

NLM and the National Institute for Dental and Craniofacial Research (NIDCR) jointly sponsored a workshop on the feasibility of creating a multimedia head and neck atlas based on Visible Human data. The participants of the workshop recommended that a head and neck atlas be produced; software tools for image segmentation and alignment be developed; and that methods be found which would reduce the histological artifacts and increase the contrast between adjacent structures in any future Visible Human data sets. Based on these recommendations NLM and NIDCR were joined by the National Institute for Deafness and other Communication Disorders, the National Eye Institute, the National Cancer Institute, the National Science Foundation, the National Institute of Mental Health, and the National Institute for Neurological Diseases and Stroke in sponsorship of this initiative.

Work is progressing on a contract that was awarded to the University of Colorado Medical Center in Denver to develop a proof of concept Web-based, multimedia interactive head and neck atlas based on current Visible Human data. In recognition of the continuing advances in network technology as well as the anticipated improvements in histological, segmentation and alignment techniques, a modular approach is being employed so that data based on such advances can be rapidly incorporated.

Work is also progressing within the research consortium that was formed through six 3-year contracts to produce a public domain software toolkit designed to focus on the "computational" problems of segmenting and aligning Visible Human and patient specific CT and MRI data. The toolkit, to be distributed by NLM in the public domain as open source

software, will feature a common API and data structure.

A third research project is designed to develop the histological methods recommended by the NLM-NIDCR workshop. Specific research problems include finding new ways to prevent brain and other tissue herniation as an artifact of freezing the specimens; establishing appropriate techniques for arterial and venous injection or another method to increase contrast with surrounding tissues; developing an appropriate technique for staining nerves during cryosectioning or another method to increase contrast with surrounding tissues so as to enable the imaging nerves at a very fine level of detail; and providing fiducial markers which would aid the alignment of the tissue images with the corresponding CT and MRI images.

Next Generation Internet

NLM is working to define Next Generation Internet (NGI) capabilities that will be required so that the NGI can be used routinely in health care, public health and health education, and biomedical, clinical and health services research. These capabilities include:

- Quality of Service
- Security and medical data privacy
- Nomadic computing
- Network management
- Infrastructure technology as a means for collaboration

NLM is currently in Phase 2 of a three phase effort to support test-bed projects that demonstrate the need for and use of NGI capabilities within the health care community. The projects being supported are to be designed to improve our understanding of the impact of NGI technology on the nation's health care, health education, and health research systems in such areas as cost, quality, usability, efficacy and security. Phase 1, resulting in 24 contracts, was a 9-month planning effort designed to identify the relevant outcomes, processes and cost variables and present a strategy for their measurement. This was followed by the current 3-year Phase 2 effort. Fifteen research contracts were awarded to support the implementation of

Phase 1 like plans within a limited geographic scope. Phase 3, a 2-year effort projected to begin next year, will test the scalability of Phase 2 projects to a national scope.

NLM sponsored a study by the Computer Science and Technology Board of the National Research Council to define the technical capabilities that the NGI must provide in order for it to support the demands of health care applications. The study, completed and

published during this past year (*Networking Health: Prescription for the Internet*), identified likely health care applications; examined their demands for such characteristics as bandwidth, quality of service, security, access, and those capabilities that are unique to health care applications as opposed those required by more general NGI based applications; and recommended an appropriate strategy for implementing these capabilities.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

David Lipman, M.D.
Director

The National Center for Biotechnology Information (NCBI), established in November 1988 by Public Law 100-607, is a division of the National Library of Medicine. The establishment of the NCBI by Congress reflected the important role information science and computer technology plays in helping to elucidate and understand the molecular processes that control health and disease. NCBI celebrated its 10th anniversary in November 1998, marking a decade of growth. Over this period of time, NCBI has established itself as a leading national resource for molecular biology information.

From its inception, NCBI has been charged with providing access to public data and analysis tools for studying molecular biology information. Over the past 12 years, the capacity to manage vast amounts of complex and diverse biological information has truly come of age, fully integrating the role of bioinformatics into the scientific process. It is now almost impossible to think of an experimental strategy in biomedicine that does not involve some online foray into the scientific databases. At the core of this shift is the recent flood of genomic data, most notably in the amount of gene sequence and mapping information. As NCBI enters into the new millennium, the horizon is a familiar one—an explosion of scientific data that must be collected, organized, stored, analyzed and disseminated. It is a great challenge for all involved—especially as the field of bioinformatics is so dynamic, and new tools and technologies that propel the field forward are continuously evolving. Through the next decade, NCBI will meet this challenge by designing, developing, disseminating, and managing the tools and technologies that will enable the gene discoveries of the 21st century.

The Center accomplishes these goals by:

- Creating automated systems for storing and analyzing knowledge about molecular biology and genetics;
- Performing research into advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules and compounds;
- Facilitating the use of databases and software by biotechnology researchers and medical care personnel; and,
- Coordinating efforts to gather biotechnology information worldwide.

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

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

GenBank—The NIH Sequence Database

The NIH GenBank DNA sequence database is an international collection of all known DNA sequences. NCBI is responsible for all phases of GenBank production, support, and distribution, including timely and accurate processing of sequence records and biological review of both new sequence entries and updates to existing entries. Integrated retrieval tools have been built to search the sequence data housed in GenBank and to link the results of a search to other related sequences, as well as to bibliographic citations. Such features allow GenBank to serve as a critical research tool in

the analysis and discovery of gene function. Interestingly, GenBank also contains a small number of sequences extracted from extinct organisms. Examples include DNA from the Neanderthal man, the woolly mammoth, the saber-toothed cat, and several giant New Zealand birds.

The total number of sequences added to GenBank reached colossal heights in FY 2000, with over four million new sequences added. The two-million mark, in terms of the total number of sequences housed in the database, was reached in January 1998; the three-million mark in December 1998; and the four-million mark in June 1999. As of October 2000, there were over nine million sequences housed in the database.

Likewise, the first billion basepairs were accumulated over a 17-year period; the second billion-basepair mark was reached in 14 months; and the third in eight months. As of October 2000, there were over 10 billion basepairs were stored in the database—a record tripling in growth. This rate of growth is far exceeding estimated growth projections. For the past several years, GenBank has experienced exponential growth with a doubling time of 16 to 18 months. Now, doubling time is less than eleven months. This indicates that, in the past six months, NCBI has processed as many bases as were processed in the previous 19 years. For the coming year, the rate of growth shows no signs of abating. The release of a “working draft” of the human genome and the conversion of the draft to a “finished” form, as well as the targeted sequencing of several other model organisms, promise that the exponential growth rate will only increase.

One of the most important sources of data for GenBank is direct sequence submission from scientists. NCBI produces GenBank from thousands of sequence records submitted directly from researchers prior to publication. Records submitted to our international collaborators, EMBL (European Molecular Biology Laboratory) at Hinxton Hall, UK and DDBJ (DNA Data Bank of Japan) at Mishima, are shared through an automated system of daily updates. Other cooperative arrangements, such as with the U.S. Patent & Trademark Office for sequences from issued patents, augment the data

collection effort and ensure the comprehensiveness of the database. Sequence data submitted in advance of publication will be maintained as confidential if requested.

When scientists submit either a DNA or amino acid sequence to GenBank, they receive an “accession number.” This number serves as a tracking device and allows the scientist to reference the sequence in a subsequent journal article. In seven years of processing direct submissions, NCBI has issued 405,000 accession numbers, with approximately 50 percent of these assigned in FY 2000. There are now 318,000 direct submission accession numbers that are publicly available and approximately 30,000 accession numbers pending release.

GenBank indexers with specialized training in molecular biology create the GenBank records and apply rigorous quality control procedures to the data. NCBI taxonomists consult on taxonomic issues, and, as a final step, senior NCBI scientists review the records for accuracy of biological information. Improving the biological accuracy of submitted data and correcting existing entries are high priorities for the GenBank team. New releases of GenBank are made every two months; daily updates are made available via the Internet and the World Wide web.

NCBI is continuously developing new tools, and enhancing existing ones, to improve access to and utility of the enormous amount of data housed in GenBank. Comprehensive coverage of all sequence data, protein as well as DNA, is provided by GenBank and the corresponding MEDLINE bibliographic information, including abstracts and publishers’ full-text documents. New in FY1999 were links to textbooks and outside sources for obtaining full-text journal articles when no direct link to the publisher was provided. The latter service, called LinkOut, also points to other external resources that may be useful in data analysis, such as biological databases and sequencing centers. With the addition of these new links, GenBank serves as a key component in an integrated database system that offers researchers the capability to perform comprehensive and seamless searching across all available data. The utility of this system is

demonstrated by the 250,000 sequence similarity and text-related searches that are performed daily.

GenBank has evolved to contain several types of DNA sequences, from relatively short Expressed Sequence Tags (ESTs) to assembled genomic sequences that are several hundred kilobases in length. EST data obtained through cDNA sequencing are critical to understanding gene function and therefore continue to be heavily represented in GenBank. As such, additional annotation is available for these sequences as part of a separate EST database (dbEST). During FY2000, NCBI continued to expand dbEST with the addition of over 3 million ESTs from several centers, including the Merck/Washington University project for humans and other organisms and the Mammalian Gene Collection project. An increasing interest in genetically modified agricultural plants and farm animals has also contributed to a total of six million sequences stored in dbEST.

Another rapidly increasing segment of GenBank is the GSS (Genome Survey Sequences) division. The GSS division of GenBank is similar to the EST division, except that its sequences are genomic in origin, rather than cDNA. Additional data on each sequence is stored in a separate database (dbGSS) and includes detailed information about the contributors, experimental conditions and genetic map locations. The dbGSS has grown nearly 84 percent during the past year and now holds 1.5 million records.

The STS (Sequence Tagged Site) division of GenBank also experienced significant growth in the past year. Sequence tagged sites are short sequences that are operationally unique in the genome and used to generate mapping reagents. A separate Sequence Tagged Sites database (dbSTS) contains additional sequence data, such as information about the contributors, experimental conditions, and genetic map locations. The dbSTS expanded to approximately 116,000 publicly available sequences in FY2000.

The “Bermuda Principle” states that all sequence data produced in publicly funded projects should be released as soon as it is “useable” for homology searching and other

types of sequence analysis. Preliminary, or “unfinished” data may be generated rapidly, but conversion to the “finished” form may take considerably longer. Hence, the impetus to release unfinished but useable data early. The NCBI, working with its international collaborators, implemented a system called the High-Throughput Genomic Sequence (HTG) division to accommodate this type of data collection. Incomplete sequences, designated as Phase 1 or Phase 2, are updated in the HTG division as work progresses and moved to the relevant organismal GenBank division upon completion. HTG sequences are substantially longer than the sequences housed in other divisions of GenBank, averaging about 114,000 basepairs. Approximately 77,000 records were added this year to the HTGS division—an increase of over 1,000 percent—reflecting the enormous growth in genome data from the Human Genome Project.

The whole genomes of over 600 organisms can now be found in Entrez Genomes. The genomes represent both completely sequenced organisms and those for which sequencing is in progress. All three main domains of life—bacteria, archaea, and eukaryote—are represented, as well as many viruses and mitochondria. New organisms added in FY2000 include *Arabidopsis thaliana* chromosome 2, *Chlamydia muridarum*, *Chlamydomonas pneumoniae* AR39, *Deinococcus radiodurans* R1, *Drosophila melanogaster*, *Neisseria meningitidis* MC58, *Ureaplasma urealyticum*, *Vibrio cholerae*, *Xylella fastidiosa*, *Bacillus halodurans* C-125, *Pseudomonas aeruginosa*, and *Buchnera sp. APS*. Sequences from these organisms will provide valuable clues for understanding the functioning of human genes. During FY 2000, the Entrez Genome division also processed complete genome sequences for 32 organelles and 84 viruses.

The Human Genome

NCBI assumed responsibility for the GenBank DNA sequence database in 1992. Under this agreement, NCBI is responsible for collecting, managing, and analyzing the growing body of human genomic data generated from the

sequencing and genome mapping initiatives of the Human Genome Project. NCBI has numerous collaborations with organizations undertaking large sequencing projects that are producing complete genome records. NCBI also plays a key role in assembling and making public the smaller genome records that can be linked and integrated with the Entrez Genomes database.

With the release of the “working draft” of the human genome, the research focus is turning from analysis of specific genes or regions to whole genomes. To accommodate this shift in research focus, NCBI has developed a suite of genomic resources to support comprehensive analysis of the human genome. Specialized tools and databases have also been designed to facilitate the use of this data.

Conversion of the human genome “working draft” and “near-finished sequences” to a “finished” version promises to be a complex task that will require the cooperation of many researchers capable of applying diverse tools to solve the problem. The data generated will more than likely be reported in a variety of forms, reflecting both the number of investigators involved and the diversity of the tools applied to solve the problem. Collection and storage of the data by individual institutions is expensive, difficult to monitor, and leads to the duplication of research efforts. All the genetic and physical maps, markers, nucleotide polymorphisms, disease phenotypes, expression profiles, and sequence data must be integrated and made uniformly accessible to the scientific community in a timely fashion if we are to reach the goal of charting and characterizing the human genome. The formulation of a common resource repository system accessible to the scientific community would serve as a nexus for the collection and storage of such diverse data.

NCBI’s Human Genome Resources web page, designed to serve as such a nexus, is linked to the GenBank sequence database and provides centralized access to a full range of genome resources, available both within and external to NCBI. This site was expanded in FY2000 to contain links to the following major resources: Map Viewer, LocusLink, UniGene, Mouse Homologies, and STS Maps. NCBI’s Human Genome Sequencing site is another

important resource supporting the human sequencing effort. This site also provides up-to-date information on sequencing efforts and access to various other types of resources such as chromosome-specific BLAST searches and data relative to specific genomic contigs.

Of particular interest to the scientific and academic communities, as well as to the lay public, is the Gene Map of the Human Genome. This website, which presents a graphical view of the available human sequence data, was originally released in October 1996 and was produced in collaboration with a team of 64 scientists worldwide. NCBI’s UniGene (Unique Human Gene Sequence Collection) was used in the compilation of the original Gene Map. In June 2000, NCBI released a large update to this site to include all the draft human sequence data. The release corresponded to the joint announcement of the completion of the working draft of the human genome. The Gene Map website integrates human sequence and map data from a variety of sources. Types of maps available include sequence, cytogenetic, genetic linkage, radiation hybrid and YAC contig maps. One can also display the location of genes, STSs and SNPs on both the draft and finished sequence. Gene Map will greatly expedite the discovery of human disease genes and is expected to result in advances in detection and treatment of common illnesses. To date, the physical map has assisted in identifying more than 100 disease-causing genes.

The Gene Map website also links to NCBI’s Genes and Disease web page, designed to educate the lay public and students on how sequencing of the human genome will lead to the identification of disease-causing genes; how these genes are inherited and cause disease; and, most importantly, how an understanding of the human genome will contribute to improving the diagnosis and treatment of disease. This site was expanded in FY2000 to include a number of additional genetic diseases and now contains descriptions for nearly 70 genetic diseases and provides links to databases and organizations that can supply additional information. For each disease-causing gene there is also a link to the PubMed literature, the Online Mendelian Inheritance in Man database (OMIM), and LocusLink.

OMIM is an electronic version of Dr. Victor McKusick's catalog of human genes and genetic disorders. The database contains 10,000 records and usage exceeds 5,800 users per day, up significantly from last year. In FY2000, OMIM was integrated into NCBI's unique search and retrieval system, which is in turn linked to several other databases. This new feature resulted in greater flexibility in field searching and increased relevance of retrieved information. New user documentation and a "frequently asked questions" page were also added in FY2000.

LocusLink, launched in FY1999, is a single-query interface to curated sequence and descriptive information about genes. LocusLink presents information on official nomenclature, aliases, sequence accession numbers, phenotypes, EC numbers, OMIM numbers, UniGene clusters, map information, and relevant web resources. LocusLink has rapidly expanded over the past year from 10,000 to 23,000 records. Major updates to LocusLink were made in FY2000 and data on the rat, mouse, zebrafish and fruitfly were added to the system. In addition, GeneRIF (Gene Reference into Function) was launched to facilitate functional annotation of loci described in LocusLink. A simple form may be downloaded that allows the scientist to provide three key pieces of information: a concise phrase describing a function or functions; a published paper describing the function; and a valid e-mail address, which remains confidential. After only 16 months, LocusLink already contains over 55,200 records.

The Reference Sequence (RefSeq) database, also launched in FY1999, provides a non-redundant set of reference standards for various naturally occurring molecules—from chromosomes to mRNAs to proteins. These standards provide a foundation for the functional annotation of the human genome and a stable reference point for mutational analysis, gene expression and polymorphism discovery. The database has grown substantially in 16 months and now holds over 20,000 nucleotide records and 35,311 corresponding protein records.

Currently, all of the sequencing work conducted at the major sequencing centers utilizes a DNA labeling chemistry combined

with an automated DNA sequencer for real-time, high-throughput data acquisition. Assembly of this data involves two steps. First, the sequencer generates a chromatograph, or trace, for each DNA sample. These traces are collected and stored in a separate computer file. The trace data is then further analyzed using a software program that interprets the trace and assigns a specific nucleotide base to a specific position, generating a complete nucleotide sequence. For many reasons, the ability of the software program to accurately interpret the trace data in a linear fashion may vary. Therefore, the accuracy of the predicted nucleotide sequence may be compromised. One way to improve nucleotide assembly accuracy is to develop an assembly algorithm that can account for the built-in error associated with the assignment of a base from the trace record to a specific position in the final sequence. NCBI investigators are attempting to do just this. A new initiative, the Genome Sequence Trace Repository, was launched in FY2000 to collect all trace data from the major sequencing centers. The data will be used to generate high-quality nucleotide assemblies and more accurate gene predictions. The data will be made available via the web to all researchers who wish to analyze the data in greater detail or generate their own assemblies independently.

The most common forms of sequence variations are single nucleotide polymorphisms, or SNPs. There has been an increasing interest in SNPs detection and discovery over the last few years as they are expected to facilitate large-scale association genetic studies. To accommodate this explosion of data, NCBI, in collaboration with the NIH National Human Genome Research Institute, launched the database of single nucleotide polymorphisms (dbSNP) in late FY1998. In its first year of public availability, dbSNP received nearly 17,000 submissions. In FY2000, 800,000 new records were added. The database was also expanded to include small-scale insertions and deletions, polymorphic repetitive elements, and microsatellite variations.

The SAGEmap database was introduced in FY1999. Serial Analysis of Gene Expression (SAGE) is an experimental technique designed to quantitatively measure gene expression. By

coupling the SAGE technique with high-throughput sequencing technology, it is possible to obtain accurate expression data for thousands of genes within a cell. A major application of SAGE is in the identification of abnormal gene expression leading to, or diagnostic of, various disease states, such as cancer.

The Cancer Genome Anatomy Project (CGAP), a collaborative project between the National Cancer Institute and NCBI, hopes to delineate the molecular fingerprints of the cancer cell. Much of the data available from the SAGE website has been generated by CGAP. The CGAP database currently contains expression data for over 20,000 human genes. The CGAP website was recently augmented with several new resources, including Gene Express—a new utility to find the computed expression level of genes in all EST libraries; the Mitelman Chromosomal Aberration Summary—a genome-wide map of chromosomal breakpoints in human cancer; and the Cancer Chromosomal Aberration Project (CCAP)—designed to expedite the definition and detailed characterization of the distinct chromosomal alterations that are associated with malignant transformation.

PubMed

PubMed is an innovative, web-based literature retrieval system that contains citations, abstracts and indexing terms for journal articles in the biomedical sciences. It also includes URLs to full-text articles from publisher websites. In early FY2000, a new version of PubMed was released that incorporated many new capabilities requested by the medical librarian community, as PubMed has now replaced the NLM mainframe search software. At this time, many functions were added or improved for limiting queries by common search filters. For example, the new version has a pull-down menu that displayed search field limits, indexes, search history, and a clipboard for gathering selected articles. Context-specific help and a “frequently asked questions” section provide guidance in making the transition to the new system.

Two major changes were also made to the links from PubMed. A component called

LinkOut was added to provide a more efficient way to store and manage the external links from PubMed and the other Entrez databases. LinkOut is a registry service that creates links from specific articles, journals, or biological data in Entrez to resources on external websites. Third parties can provide a URL, resource name, brief description of their web site, and specification of the NCBI data from which they would like to establish links.

A new PubMed tutorial was produced and added to the website in early FY2000. Additional system enhancements were made to PubMed throughout the year, including the addition of a preview capability to the history screen, a sorting feature that allows users to sort citations by author, journal or publication date, and a new toxicology subset. The Cubby service, a relatively new feature of PubMed, provides users with a Stored Search feature to store and update searches. It also allows users to customize their LinkOut display to include or exclude links to providers. The PubMed Journal Browser allows you to look up journal names, MEDLINE abbreviations, or ISSN numbers for journals that are included in the PubMed system. A list of journals with links to full-text websites is also available.

PubMed services have expanded in all aspects. Full-text journals that link to PubMed have nearly tripled, from 444 in September 1999 to 1138 in August 2000. Usage of PubMed by the scientific and lay communities has also grown considerably since its introduction in 1997. Currently, approximately 20 million searches are conducted per month and as many as 140,000 different users seek information daily via PubMed.

In collaboration with book publishers, the NCBI is also adapting textbooks for the web and linking them to PubMed. The idea is that the textbook will serve to provide accessible background material that users can explore to understand unfamiliar concepts found in a PubMed search results. The textbook, *Molecular Biology of the Cell*, 3rd ed., by Alberts et al. is the first book to be included online. Nine other textbooks are pending, and a few other texts are under discussion. In addition, the NCBI is collaborating with the NCI to include an oncology textbook.

The overall success of PubMed has led to collaboration between NCBI and the NIH Director's Office to establish a web-based repository for barrier-free access to primary reports in the life sciences. This repository, called PubMed Central, is based on a natural integration with the existing PubMed biomedical literature database. PubMed Central was launched in FY2000 with sample issues from the *Proceedings of the National Academy of Sciences* and from *Molecular Biology of the Cell*. Five journals are currently available on PubMed Central, and 48 are forthcoming. Articles may be viewed as HTML through a web browser or downloaded in PDF (Portable Document Format). Features include: links from article reference citations to PubMed abstracts, figures sized for on-screen viewing, and support for supplementary information such as data tables, streaming video, and high-resolution images.

PubMed Central serves as a nexus for scientific publishers, professional societies, and other groups with an interest in the life sciences to archive, organize, and distribute their research articles at no cost to the user. The core concept of the proposal is to remove access barriers to the scientific literature and to make it available worldwide to the scholarly community.

The BLAST Suite of Programs

Comparison, whether of morphological features or protein sequences, lies at the heart of biology. The introduction of BLAST in 1990 made it easier to rapidly scan huge sequence databases for overt homologies and to statistically evaluate the resulting matches. The journal article describing the original algorithm used in BLAST has since become the most heavily cited paper of the decade, with over 10,000 citations. BLAST compares a user's unknown sequence against the database of all known sequences to determine likely matches. Sequence similarities found by BLAST have been critical in several gene discoveries. Hundreds of major sequencing centers and research institutions around the country use this software to transmit a query sequence from their local computer to a BLAST server at the NCBI via the Internet. In a matter of seconds, the

BLAST server compares the user's sequence with up to a million known sequences and determines the closest matches.

Not all significant homologies are overt, however. Some of the most interesting are subtle and do not rise to statistical significance during a standard BLAST search. NCBI has extended the statistical methodology in BLAST to address the problem of detecting weak, yet significant sequence similarities. The so-called Position-Specific Iterated BLAST (PSI-BLAST) searches sequence databases with a profile constructed using BLAST alignments, from which it constructs a position-specific score matrix. For protein analysis, the new Pattern Hit Initiated BLAST, or PHI-BLAST, serves to complement the profile-based searching that was previously introduced with PSI-BLAST. PHI-BLAST further incorporates hypotheses as to the biological function of a query sequence and restricts the analysis to a set of protein sequences that are already known to contain a specific pattern or motif. Other new features added to the BLAST suite of programs include:

- Tax-BLAST, which allows users to limit a BLAST query to a specific organism or taxonomic group;
- Reverse PSI-BLAST (RPS-BLAST), which is used to identify conserved domains in a protein query sequence;
- BLAST 2 Sequence, a tool for aligning two nucleotide or protein sequences—producing a pairwise DNA-DNA or protein-protein sequence comparison;
- IgBLAST, which is used to facilitate analysis of immunoglobulin sequences;
- IMPALA, software that matches a protein sequence against a library of score matrices stored from PSI-BLAST; and
- The Conserved Domain Search Service (CD-Search), which can be used to identify the conserved domains present in a protein sequence.

An extensive BLAST “information guide” was also introduced in FY2000 that provides query, BLAST, and PSI-tutorials.

The BLAST sequence searching server is one of NCBI's most heavily used services and its usage continues to grow at a pace reflecting

the growth of GenBank. Each day more than 70,000 sequence searches are performed, with users submitting their requests through e-mail, server/client programs, and the World Wide Web. The popularity of BLAST has stressed the existing computing capacity and additional computing resources have been added to accommodate the growing volume of users and expansion of the sequence databases. Therefore, a new system called QBLAST was developed to better handle the increasing BLAST load. This system obviates the need for persistent connections while users are waiting for results and allows NCBI to better distribute the query load.

Other Specialized Databases and Tools

Several specialized web services were released or substantially updated throughout the year. The recently announced Conserved Domain Database (CDD) is a collection of sequence alignments and profiles representing protein domains conserved in molecular evolution. It includes domains from *Smart* and *Pfam* (two popular web-based tools for studying sequence domains), as well as domains contributed by NCBI researchers. CDD can be used to identify conserved domains in a protein query sequence, using the recently released CD-Search. CD-Search uses RPS-BLAST to compare a query sequence against position-specific score matrices that have been prepared from conserved domain alignments present in the CDD. Hits can be displayed as a pairwise alignment of the query sequence with a representative domain sequence, or as a multiple alignment. Alignments are also mapped to known 3-dimensional structures, and can be displayed using Cn3D, also recently revised. In the Cn3D display, residues in sequence alignments are variously colored, based on their degree of conservation. New features in Cn3D 3.0 include advanced graphics, improved sequence and alignment viewers, and coloring by sequence alignment conservation.

NCBI issued a new release of the Clusters of Orthologous Groups (COGs) database that provides new functionalities and improved genome annotation. The COGs database is a natural system of gene families

from complete genomes. COGs were identified using a technique that compared the protein sequences encoded in 21 complete genomes representing 17 major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogs from at least three lineages and thus corresponds to an ancient conserved domain. In addition to phylogenetic patterns, the COGs may now be searched using free-text words or protein and gene names using the search tool COGnitor. Analysis of COGs shows the molecular similarities and differences between species, which not only can provide clues about evolution, but also may help to identify protein families, predict new protein functions, and point to potential drug targets in pathogenic species.

The Eukaryotic Organelles Home Page, released in FY2000, provides an overview of eukaryotic organelles; a description of the Organelle Reference Sequences project (part of the RefSeq database described previously); and links to (a) lists of completely sequenced organelles shown in taxonomic hierarchy and alphabetically by organism, (b) gene and RNA order in metazoan mitochondria, and (c) related web sites. For example, by linking to the Entrez Genomes website, one can view a graphical representation of a specific eukaryotic organelle. Organelles may be viewed in their entirety or explored on a smaller scale in progressively greater detail. This site also displays associated sequence data and provides a summary of Coding Regions for each organelle.

The Gene Expression Omnibus (GEO), developed in FY2000, is a gene expression data repository and online resource for the retrieval of gene expression data from any organism or artificial source. Many types of gene expression data from different platform types are accepted, accessioned, and archived as a public data set, including spotted microarray, high-density oligonucleotide array, hybridization filter and serial analysis of gene expression (SAGE) data. A series of precomputed definitions and descriptions of the data, as well as online tools for the interactive retrieval and analysis of this expression data, are currently under development.

Genomic information for organism-specific resources was greatly augmented

throughout the year. Web pages for the fruitfly, mouse, rat, and zebrafish were added. Each site contains descriptive information, precomputed BLAST search results, built-in query boxes for searches conducted against other NCBI databases, and links to many other search tools. These sites are complemented by the addition of a new resource called HomoloGene. HomoloGene is a tool that compares nucleotide sequences between pairs of organisms—including human, mouse, rat, zebrafish, and fruitfly—in order to identify putative orthologs. Curated orthologs are incorporated from a variety of sources via LocusLink.

Work continued throughout the year on the Molecular Modeling Database (MMDB), a database of three-dimensional biomolecular structures derived from X-ray crystallography and NMR-spectroscopy. MMDB currently houses nearly 12,000 structures. MMDB is a subset of three-dimensional structures obtained from the Brookhaven Protein DataBank (PDB), excluding theoretical models. MMDB reorganizes and validates the information in a way that enables cross-referencing between the chemistry and the three-dimensional structure of macromolecules. By integrating chemical, sequence, and structure information, MMDB is designed to serve as a resource for structure-based homology modeling and protein structure prediction.

The Mammalian Gene Collection (MGC) is a new effort sponsored by the NIH to generate full-length complementary DNA (cDNA) resources. This project will make all of the cDNA resources generated accessible to the entire biomedical research community. The MGC project involves the production of cDNA libraries and sequences, database and repository development, and support of research efforts leading to improved library construction, sequencing, and analytic technologies.

The PROW (Protein Reviews On the Web) site was also redesigned this year and a new search engine was installed, improving overall search capacity. PROW is an online resource that features PROW Guides—authoritative, short, structured reviews on proteins and protein families. The Guides provide approximately 20 standardized categories of information, such as abstract,

biochemical function, ligands, and references for each available protein.

The purpose of the NCBI Taxonomy Project is to build a consistent phylogenetic taxonomy for the NCBI sequence databases. During FY2000, members of the taxonomy group maintained the overall structure of the taxonomy database and web pages, monitored the literature for new insights, and maintained contact with off-site taxonomy advisors. NCBI taxonomists also provided consultation to staff of the EMBL Data Library and DNA Database of Japan, collaborating sequence databases in Europe and Japan. Members continued to add new species or perform other edits to the database as required. Members also guided the NCBI indexing staff on taxonomic issues.

The Taxonomy Database, one component of the taxonomy project, provides general information on taxonomic resources as well as a list of outside curators currently collaborating with NCBI taxonomists. The database contains the names and lineages of the greater than 85,000 organisms represented by at least one nucleotide or protein sequence in the NCBI genetic databases. The database is recognized as the standard reference by the international sequence database collaboration. The Taxonomy Browser is an NCBI-derived search tool that allows an individual to search the database. Using the browser, information may be retrieved on available nucleotide, protein, and structure records for a particular species or higher taxon. Over 25,000 names were added to the browser this year.

The NCBI also introduced in FY2000 a password-protected, in-house database called the abdb, or Aberration Database. This database contains chromosome structures in cancer and chromosomal comparative genome hybridization data.

FY2000 was particularly productive in terms of designing and developing new tools to display and analyze sequences contained in the NCBI databases. New additions to the NCBI website include BankIt, CD-Search, Entrez Map Viewer, HomoloGene, Sequin, UniGene, and VecScreen. In addition, new features were added to the many existing web-based tools in order to increase their value to the research community and to improve and simplify both the data

submission and search processes. Some of these have already been discussed. Those that have not are discussed below.

A new version of BankIt—a sequence submission software package—is now available. BankIt 3.0 provides submitters with two new features to allow more specific sequence annotation. The source organism modifier list has been expanded and all coding regions will be conceptually translated. Sequin, another sequence submission software program, was also recently updated. This software is designed to simplify the sequence submission process by providing graphical viewing and editing options along with validation checks of the data.

NCBI updated the design of the ORF Finder, a graphical analysis tool that finds all open reading frames of a selected minimum size in a user's sequence or in a sequence already in the database. ORF Finder identifies open reading frames using the standard or alternative genetic codes. The deduced amino acid sequence can be saved in various formats and searched against the sequence database using the WWW BLAST server. A new feature allows the user to perform a comparison of candidate protein COGS using the COGnitor tool.

A new version of Map Viewer allows visualizing and integrating genome sequence with a variety of maps. The Map Viewer, a software component of Entrez Genomes, displays one or more maps that have been aligned to each other based on shared marker and gene names, and, for the sequence maps, based on a common sequence coordinate system.

UniGene (Unique Human Gene Sequence Collection) is an experimental system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters. Each UniGene cluster contains sequences that represent a unique gene, as well as related information such as the tissue types in which the gene has been expressed and map location. In addition to sequences of well-characterized genes, hundreds of thousands novel expressed sequence tag (EST) sequences have been included. UniGene now contains 82,500 human clusters, 69,000 mouse clusters and 37,000 rat clusters. Sequencing centers are now developing unique markers based on UniGene clusters, and mapping these markers on

radiation hybrid panels. NCBI has developed algorithms (CONCORD) for producing higher density radiation hybrid maps. Higher density maps are much more accurate and allow for increased precision in identifying genes associated with radiation hybrid markers. Consequently, UniGene has become a critical resource for the sequencing community.

VecScreen is a new tool for screening a nucleotide sequence for vector, linker, and adapter contamination. This tool will help researchers identify and remove any segments of vector origin prior to sequence analysis or submission. NCBI developed VecScreen to minimize the incidence and impact of vector contamination in public sequence databases.

Collaborations with other NIH institutes continue to flourish. For example, the Malaria Genetics and Genomics web site was developed in collaboration with the National Institute of Allergy and Infectious Diseases. These resources include organism-specific sequence BLAST databases (*Plasmodium falciparum* only, all *Plasmodium*, and all *Toxoplasma*), genome maps, linkage markers, and information about genetic studies. Links are provided for other malaria web sites and genetic data on related apicomplexan parasites, including *Toxoplasma gondii*. Additional resources will be added as they become available. Numerous additions were made to this web page in FY2000 and include maps, markers, recombination data, crossover counts/locations, and genotype segregation proportions for the linkage groups corresponding to *Plasmodium Falciparum* chromosomes. Data from *P. berghei* and *P. vivax* projects was also added to BLAST via the Malaria CustomBLAST, and a separate *P. vivax* Division web page was constructed to provide species-specific information for this organism, the second most prevalent form of human malaria.

Database Access

Entrez Retrieval System

The major database retrieval system at NCBI, Entrez, was originally developed for searching nucleotide and protein sequence databases and related MEDLINE citations. It

was later expanded to include the integrated set of PubMed, MMDB (Molecular Modeling Database) 3-D Structure, Genomes, and Taxonomy databases. Users can search gigabytes of sequence and literature data with techniques that are fast and easy to use. A key feature of the system is the concept of “neighboring,” which permits a user to locate related references or sequences by asking for all papers or sequences that resemble a given paper or sequence. The ability to traverse the literature and molecular sequences via neighbors and links provides a very powerful and intuitive way of accessing the data. Entrez users submit 900,000 text searches and 70,000 sequence similarity searches daily. Over 180,000 Entrez DNA and protein queries per weekday are handled and the number continues to increase.

Entrez’s design permits incorporating additional linked databases without changes in the user interface. Web Entrez now provides graphical views of nucleotide and protein sequences and access to the NCBI Genomes database, which contains additional graphical views of sequences and chromosome maps. The structure viewer, Cn3D, permits visualization of 3-dimensional protein structures and offers a greatly expanded array of annotation tools, including the ability to define molecular features and specify their display characteristics, global save capabilities, and an improved installation process.

During FY2000, Entrez was overhauled to accommodate additional data resources and to facilitate the more finely tuned searches demanded by the explosion of data in the sequence, structure and literature databases. The new Entrez system not only offers an enhanced search interface to the five familiar databases—Nucleotide, Protein, Structure, Genome, and PubMed—but also adds a sixth database to the mix called PopSet, short for population studies. This new release has a completely new back-end, more databases, a new advanced search screen that records a history of all completed queries, and a clipboard to save articles on interest for later use. This system also has full support for LinkOut, a registry service designed to create links from specific articles, journals or biological data to resources on external websites. A new Entrez PubMed tutorial was developed

and added to the website in early FY2000. Several other specialized Entrez search and retrieval system tutorials have also been developed, including tutorials for the Nucleotide, OMIM and Structure databases.

Other Network Services

Usage of NCBI’s web services, first introduced in December 1993, continues to expand as more services are added. NCBI staff continued to make access and usage easier with improved documentation and tutorials. General information about NCBI, its databases and services, data submissions and updates, and NCBI investigator projects, as well as an ever-increasing number of search tools, are readily available via the web. The web server also provides capabilities for Entrez and BLAST searches and data submission through BankIt. Many other web servers have links to the NCBI server in order to conduct searches and obtain the latest GenBank records. At the end of FY2000, NCBI’s site was averaging over 9,000,000 hits daily. Because of the mission-critical nature of NCBI’s computing platforms for PubMed, Entrez, BLAST, and other services, an extensive program in system monitoring has been implemented. Based on measurements taken every 15 minutes from 50 sites across the U.S. and overseas, the average time to load the entire NCBI home page is now under 1.5 seconds, an average PubMed search takes less than 3 seconds and availability has been better than 99 percent.

GenBank is also distributed over the Internet through the standard File Transfer Protocol (FTP) program, and many large commercial and academic sites maintain a local copy of GenBank. NCBI’s Data Repository, with over 50 additional molecular biology databases, is also distributed via FTP. Over 1,000 sites download greater than 180 gigabytes of data daily, including daily GenBank and dbEST update files. There are 5,000 FTP requests per day.

NCBI maintains two electronic mail servers, BLAST and QUERY. The BLAST server performs sequence similarity searches and QUERY retrieves records from several sequence databases, including GenBank, EMBL, Swiss-

Prot, and PIR. Any user in the world with e-mail access can submit a query to either server and have a reply within minutes. More than 11,000 queries are handled daily by the BLAST and QUERY servers.

The improvement of NCBI's sequence submission software continued to be a high priority. A new version of Sequin, NCBI's stand-alone submission tool, was released in FY2000. Sequin now allows the author to edit complete bacterial chromosomes or large eukaryotic chromosomal segments in a single record. In addition, Sequin can now function as a stand-alone or network-aware program. Complete documentation and a tutorial are maintained on the web site. The web submission tool, BankIt, was also revised. The new version automatically checks a sequence to ensure that it is complete and free of contamination. BankIt is now in its sixth year of use. With the addition of Sequin, submissions entering via the BankIt route have dropped from about 50 percent to just over 30 percent in the final quarter of FY2000.

During FY2000, NCBI upgraded a number of its key systems to keep pace with the increase in demand for public services, such as BLAST and PubMed, as well as to accommodate the dramatic increase in the growth of GenBank. This increase in demand, as well as the growth of GenBank, was a direct result of the release of the "working draft" of the human genome. To handle the influx of new data, NCBI acquired additional storage and computing capacity and upgraded the network core infrastructure.

NCBI began building a "compute farm" based on Intel 8-way processors running the Solaris operating system and using a commercial queuing system for batch job submission and load balancing. The farm currently contains 88 processors. The queuing software enables NCBI to create and manage logical clusters devoted to specific functions, for example, BLAST and PubMed neighboring. Each cluster can be scheduled and prioritized independently, thereby maximizing the utilization of the cluster. Furthermore, idle cycles on server or workstations can be transparently incorporated into the queues for even further efficiency. A recently installed high-performance, shared

network file server provides large capacity data storage.

As the installation of a high-performance network server would place too large a burden on the local area network, NCBI upgraded the core network infrastructure to multigigabit speed. Two Cisco Catalyst 6506 routing switches were installed and connected via a 2-gigabit link. Each unit has 50 ports for GigabitEthernet connections to servers and other network devices, which in turn, provide 100-megabit connections to desktop systems. The routing capacity of each routing switch is approximately 10 times that of the single router they replaced.

The expansion of available network bandwidth made possible another major computing initiative. A pilot project was launched to investigate the feasibility of replacing expensive and management-intensive Unix workstations with windowing terminals connected to Unix servers. The terminals, called SunRays, have proven successful, providing a user experience that is essentially indistinguishable from a dedicated workstation and at a fraction of the cost and administrative overhead. This project will be greatly expanded in the coming fiscal year.

In addition to these major projects, a number of important smaller additions were made to the existing computing infrastructure, including:

- The two primary GenBank database servers were upgraded from two-processor systems to four-processor systems, with additional disk and memory;
- Approximately 400 GB of mirrored storage was added to the primary PubMed database servers;
- Three new 4-CPU PubMed web servers with associated storage were added;
- The existing public FTP service was replaced with a dedicated highly-available cluster comprised of two 2-CPU systems;
- Approximately 900 GB was added to an existing Compaq Storage Area Network that provides storage for three back-end Sybase database servers;

- A new 4-drive/60-slot AIT/2 tape library was installed to provide additional network backup capacity;
- Three new 8-CPU Dell servers were installed to augment the public BLAST search service;
- Memory and disk storage were added to numerous existing servers;
- Three Windows NT servers were replaced with new models with more than triple the CPU throughput, memory and disk;
- A single-CPU Linux compile server was replaced with a new 2-CPU system; and
- A 2-CPU BLAST development server was installed.

Equally important as building databases for molecular sequence information is the ability to access and retrieve the information using automated systems. The NCBI software toolkit concept addresses this need by creating software modules that provide a set of high-level functions to assist developers in building application software. Among these tools are a Portable Core Library of functions in the C language that facilitate writing software for different hardware platforms and operating systems, and AsnLib, a collection of routines for handling ASN.1 data and developing ASN.1 applications. The ASN.1 (Abstract Syntax Notation) tool is an International Standards Organization data description language that provides a mechanism for defining and structuring data as well as a set of program definitions that can interact with databases structured in ASN.1. With ASN.1 definitions and the NCBI software toolkit, complicated analysis programs can be readily constructed from pre-existing sets of modular tools, saving considerable time and programming effort.

Basic Research

Basic research is at the core of NCBI's mission. The Computational Biology and Information Engineering Branches are the main research branches of NCBI. Each Branch is comprised of a multidisciplinary team of scientists that carries out research on fundamental molecular biomedical questions by

developing and applying mathematical, statistical and other computational methods to the life sciences. The research approach taken relies on both the theoretical and applied sciences, as, in the field of bioinformatics, these two lines of research prove mutually reinforcing and complementary. Research conducted by NCBI investigators has led to development of many new theoretical and practical models and application of these methods to the life sciences has opened the doors to new areas of research. For example, the development and application of novel or improved algorithms to biologically important molecules has led to the identification of many previously unknown molecular structures. Structure identification, in turn, provides important clues as to how a molecule functions. Once you have an understanding of molecular function, you can begin to elucidate its natural role in a particular molecular pathway. From here, you can study what happens in a diseased state.

More specifically, investigators have developed novel protein sequence search methods using "patterns as seeds"; developed a pairwise sequence comparison method to assess a possible relationship between two sequences; developed novel database designs, data management and analysis techniques, such as for single nucleotide polymorphism (SNP) data; developed measures of threading statistics to identify structure relationships; and designed methods to analyze the interactions among quantitative traits in the course of sympatric speciation. Macromolecular sequence analysis programs have been applied to investigate chaperone-like ATPases; POZ domains; HD domains; "death" domains in apoptosis; DNA polymerase beta-like domains in nucleotidyltransferases; GPS domains in polycystin-1; START domains in signaling proteins; selenium-containing proteins; HIV-1 haplotypes and HTLV-I subtypes; and AT-hook motif containing proteins. In addition, many other families of proteins from previously published works are being further studied with the goal of identifying new protein family members, especially in fully sequenced genomes.

Genome-scale projects have continued to be a staple of Computational Biology Branch

research. Projects have included genome analyses using clusters of orthologous groups of protein sequences from whole proteomes (all of the proteins found in cells and tissues); the distribution of protein folds in the three superkingdoms of life; the comparison of complete proteomes of *Caenorhabditis elegans* and *Saccharomyces cerevisiae*; the evolution of protein families in Archaea; finding systematic annotation errors in whole genome annotations; the use of metabolic pathways for functional annotation of genome sequences; the analysis of regulatory sequence elements in cell cycle-dependent transcription; the analysis of the genome of the human pathogen, *Chlamydia trachomatis*; and the analysis of the genome of the malaria parasite, *Plasmodium falciparum*.

Other research efforts include the further development of a database of histones and histone fold sequences and structures; research on a model of HIV infectivity; construction of a genetic linkage map of microsatellites in the domestic cat; analyses of the horizontal transfer of genes using the evolution of aminoacyl-tRNA synthetases as an example; and the prediction of curved DNA in promoter sequences.

Currently, the intramural group is engaged in over 20 projects, many of which involve collaborations with other NIH institutes as well as with academia and private industry. A Board of Scientific Counselors, comprised of extramural scientists, meets twice a year to review the research activities of the Center (see Appendix 4 for list of members). The high caliber of the work is evidenced by the number of peer-reviewed publications, well over 100 in FY2000, as well as by the number of outside collaborations. The staff participated in over 80 oral and 28 poster presentations at major scientific meetings; 79 presentations to academic departments and companies engaged in molecular biology research; 53 presentations to visiting delegations oversight groups and steering committees; and 19 presentations at NCBI's computational biology lecture series. This year NCBI initiated a new lecture series on mathematical topics in the biological sciences that involved six staff presentations. NCBI also hosted presentations by 46 invited speakers.

The Visitors' Program continues to be successful in recruiting members of the external

scientific community to engage in collaborative research with members of the NCBI Computational Biology Branch. Members of the Visitors' Program also participated in joint activities of database design and implementation with the Information Engineering Branch. The Visitors' Program, administered in conjunction with Oak Ridge Associated Universities, facilitated 55 visits by 53 individual researchers, including 2 long-term sabbaticals, this past year.

The NCBI GenBank Post-Doctoral Fellow program, designed to provide for concentrated efforts on improving and strengthening GenBank, is currently filled. The NCBI uses the NIH Intramural Research Training Award Program and the Fogarty Visiting Fellow mechanisms to recruit for this program.

User Support

As part of its mandate to facilitate the use of databases and software by the scientific community, NCBI maintains a user support group staffed by scientists, librarians, and information specialists with broad experience in interpreting both basic biology and medical information. The focus of this group is to support the varied services that NCBI offers. To accomplish this, NCBI support staff are prepared to address all questions from users and may be contacted by e-mail, phone, or fax. NCBI has recently extended its outreach programs to the library science community and has sponsored both oral presentations and workshops on various biotechnology information topics.

The number of NCBI services offered to the scientific community continues to grow, as does the number of daily users, leading to an increased need for support services. In addition, as the scope of the services offered by NCBI expands, so must the scope of user support system. The NCBI Information Resources Branch staff, with the support of contract personnel, provides a timely response to all user queries. The three main areas of user support include: general information regarding GenBank and related database services, as well as general information on data submission; technical assistance for submission of new GenBank data and revisions or updates to existing records; and

technical assistance with Entrez and other data retrieval systems. Scientists who submit a sequence to GenBank are furnished with an "accession number" within 48 hours. This number serves as a tracking device and allows the scientist to reference the sequence in a journal article as well as to support staff when requesting assistance in revising or updating a submission. Most responses to user inquiries are immediate. A response to a request that requires the input of additional expertise is usually provided within 24 hours. In FY2000, staff from NCBI, the MEDLARS Management Section, and the Reference Section provided assistance with general PubMed inquiries.

Outreach and Education

In FY2000, NCBI expanded its outreach and education programs to increase awareness of its myriad of public databases and specialized tools and services. NCBI staff presented at numerous scientific exhibits, seminars and workshops; sponsored a number of training courses—both lecture courses and "hands-on" courses; and published and distributed various forms of printed information.

Staff participated in 14 national exhibits and presented six training courses or lectures at five professional society meetings. Numerous researchers served as faculty at seven workshops sponsored by external institutions, including Cold Spring Harbor, Woods Hole, and the Smithsonian. In cooperation with various academic institutions, NCBI staff presented nine informational courses for research scientists and five courses for librarian/information specialists. NCBI staff also presented five training courses held at various other governmental sites.

NCBI sponsored a lecture series on the NIH campus for both NIH and local-area scientists. This series was offered five different times and included both lecture and hands-on sessions. NCBI also offered for the first time two mini-courses for NIH scientists. The first course, "Making Sense of DNA and Protein Sequences," was offered three times. In order to accommodate the busy schedules of the NIH scientists, the course could be completed either in a classroom setting, at the lab bench via the web, or downloaded to a user's local

workstation and completed at their leisure. In the latter capacity, the course served as a protein analysis tool that could be applied to the user's research. The second mini-course focused on Entrez and PubMed and was offered two different times in a classroom setting. NCBI is currently working towards making this series available via the web.

In FY2000, the NCBI, in collaboration with the NIH Clinical Center Library, offered individualized computer demonstrations for NIH staff. NIH staff members could also schedule a one-on-one consultation with an NCBI staff member as to how they could apply an NCBI service to their research interest. Nine scientists signed up for the demonstration program and 41 scientists scheduled a consultation.

NCBI News is a quarterly newsletter that informs the scientific community about NCBI's current research activities, as well as the availability of new database and software services. The newsletter contains information on user services; announcements of new or updated tutorials; a section on frequently asked questions; NCBI investigator profiles; and a bibliography of recent staff publications. In FY2000, over 23,000 printed copies of the *NCBI News* were distributed quarterly. The newsletter is also available to the general public via the NCBI website.

NCBI has developed a comprehensive fact sheet that outlines the services and databases offered by NCBI and highlights where to find them on the World Wide Web. The NCBI also develops and distributes fact sheets that focus on a particular service or database. In addition, a number of other informational and educational resources are available on the NCBI website. "Articles of Interest" provides the user with a brief introduction to the field of bioinformatics and links to articles describing different NCBI resources. Another link discusses the fundamental principles underlying sequence similarity search tools. Interactive tutorials may also be found for a number of databases and search and retrieval tools. For example, "How to BLAST" is an interactive tutorial designed to help the first-time BLAST user employ this tool in their research. Tutorials for Entrez, PubMed and OMIM were recently

revised to incorporate the many new features added to these systems during the past year.

“Coffee Break,” a new resource at NCBI, is a collection of short reports on recent biological discoveries. Each report incorporates interactive tutorials demonstrating how bioinformatics tools are used as part of the research process. Each report is approximately 400 words and is usually based on a discovery reported in one or more recent articles from the peer-reviewed literature. The topics change every few months and public suggestions for future topics may be submitted to NCBI directly through this site.

NCBI in the News is a selective, annotated compilation of articles that reference NCBI programs or staff members and includes articles from the mass media as well as from the scientific and technical publications. In FY2000, NCBI was referenced in over 100 articles.

Extramural Programs

Funding for extramural bioinformatics activities is the responsibility of the Library’s Extramural Programs. NLM funds research projects in areas defined as important to its mission. As the nation’s premier repository of biomedical information, NLM has a vital interest in information management and in the enormous utility of computers and telecommunication for improving storage, retrieval, access, and use of biomedical information. In this context, a wide variety of research in computational biology has been supported through the program, including methods and algorithms for sequence analysis, structure and function prediction, new machine architectures and specialized databases. Extramural postdoctoral training in the cross-

disciplinary areas of biology, medicine, and computer science is also funded through the NLM informatics fellowship program.

Biotechnology Information in the Future

In the past year, there has been an explosion in the volume of genomic data produced by the scientific community, most notably in the amount of gene sequence and mapping information. This is due in a large part to the recent release of the first draft of the human genetic code. The commitment to providing the scientific community with the fully deciphered genetic code as quickly as possible, as well as recent advances in molecular analysis technologies, promises that the exponential growth in genomic data will only increase. This reinforces the need to build and maintain a strong infrastructure of information support. NCBI, a leader in the fields of computational biology and bioinformatics, will play an active and collaborative role in deciphering the human genome and in developing of state-of-the-art software and databases for the storage, analysis, and dissemination of data. The genomic information resources developed and disseminated thus far by NCBI investigators have contributed significantly to the advancement of the basic sciences and serve as a wellspring of new methods and approaches for applied research activities. The value of these resources will continue to grow, as NCBI is committed to the challenge of designing, developing, disseminating and managing the tools and technologies enabling the gene discoveries that will significantly impact health in the 21st century.

EXTRAMURAL PROGRAMS

Milton Corn, M.D.
Associate Director

The Extramural Programs Division (EP) of NLM continues to receive its budget under two different authorizing acts: the Medical Library Assistance Act (unique to NLM), and Public Health Law 301 (covers all of NIH). The funds are expended mainly as grants-in-aid, and in some instances as contracts, to the extramural community in support of the goals of the Library. Review and award procedures conform to NIH policies. A list of awarded grants is on the Extramural Programs web site.

EP issues grants in a broad variety of programs, all of which pertain to informatics and information management with the exception of the Publications Grant program.

- Resource Grants for information management; usually involve medical libraries
- Training and fellowship grants in support of informatics research training
- Research Grants in informatics, information science, and biomedical computing
- Research Resource grants to support informatics and bioinformatics research
- Publication grants to support preparation of scholarly manuscripts
- SBIR/STTR
- Special Projects

Resource Grants (MLAA)

Resource Grants, authorized by the Medical Library Assistance Act, support access to information as well as promote networking, integrating, and connecting computer and communications systems. There are four types of Resource Grants which range in complexity as well as in dollar amounts and duration. They are considered “seed” grants designed to initiate a resource or service or program that is expected to become self-sustaining. All four Resource Grants are open to public and private, nonprofit health institutions/organizations engaged in health education, research, patient care, and

administration are eligible, and all four strongly encourage some health science library involvement in the project.

Information Access Grants

Information Access Grants, aimed primarily at hospitals, clinics, community health centers and similar small health organizations, support installation of computers and other information technology as well as training to facilitate access to NLM’s PubMed and other databases and/or improve efficient distribution of the library resources within a region. These grants provide up to \$12,000 per participating institution and are available to single as well as multiple institutions working together.

Information Systems Grants

Information Systems Grants, ranging up to \$150,000 per year for up to three years, are intended for more complex projects and organizations than are the Access grants, and are suitable for a broad variety of information management projects at larger hospitals, medical schools and other health-care related institutions. These grants can be used to support both personnel and information technology, and have been widely useful in a number of areas. Planning grants are also available for those who are not quite ready to request the standard Information Systems grant.

Internet Connections Grants

The Internet Connection Grants provide grants up to \$30,000 to single institutions and up to \$50,000 to multi-institution conglomerates to initiate Internet access. Funds are usually used to pay for gateway/router equipment, Internet Service Provider fees, and line charges in the first year. Some institutions with existing Internet access can use these grants to improve distribution of Internet access internally, or to extend access to other institutions.

Interest in these grants diminished somewhat during the middle 90s but has been steady in recent years at a level of \$400,000-500,000 per year. A recent study commissioned by Extramural Programs suggests that the

program would benefit from a simpler application form, more suitable for use by inexperienced applicants, and by more publicity to community health care organizations offices.

IAIMS Grants

Integrated Advanced Information Management Systems (IAIMS) Grants are designed to facilitate institution-wide information systems that link individual and organizational databases and information systems for patient care, education, research, library, and administration. IAIMS Grants support two phases, planning and implementation, with the goal being to support organizational mechanisms that foster the integration of various information systems, and the organization's short- and long-term planning for optimal use of information technology. The planning phase funds up to \$150,000 for one to two years; the operational phase funds up to \$500,000 per year for five years or \$550,000 with an IAIMS apprenticeship option.

Although the program was initially intended to support a minimal set of models that could then serve as templates for others, experience with the grants demonstrated that the problems and therefore the solutions were parochial. It became clear that an IAIMS climate required much more emphasis on people and organizational issues than on information technology, which meant that a chief value of the grant was for smoothing the managerial interactions essential to the IAIMS goals. Although the large increase in dollar outlay for information technology by medical centers in recent years dwarfs the value of the grants, interest remains high, perhaps because of their impact on otherwise stubborn management and organizational problems.

Training And Fellowships (MLAA)

Exploiting the potential of computers and telecommunication for health care information requires investigators who understand biomedicine as well as fundamental problems of knowledge representation, decision support, and human-computer interface. NLM remains the principal support for research

training in medical informatics, including clinical and basic science domains. NLM provides both institutional and individual mechanisms of support for its training activities.

NLM-Supported Training Programs

Five-year institutional training grants support approximately 150 trainees at pre-doctoral and postdoctoral levels. Twelve institutions currently receive such support, but because a number of these share support with other universities and teaching hospitals, there are over 20 training sites. For the past few years, NCI and NIDR contributed funds to NLM to help support slots at these training sites for applicants interested in cancer radiation therapy and dental informatics respectively. NCI discontinued its support in FY 1999.

BISTI and NLM's Training Programs

Interest in biomedical computing exploded at NIH after the June 1999 publication of the Biomedical Information Science Technology Initiative (BISTI) report on biomedical computing. Because the BISTI report stimulated a new set of pan-NIH grant programs to begin for the most part in FY 2001, NLM provided each of the 12 institutional training programs with "BISTI" administrative supplements of \$200,000 during FY 2000 as a means of initiating or enhancing training tracks in bioinformatics. The BISTI report recognized that biomedical researchers need much more training in the tools of informatics, but in addition there is now and will continue to be a marked need for informaticians with sufficient domain knowledge to develop these tools as data and data interpretation become increasingly complex. Because of its long history of supporting informatics research training, NLM is well-poised to make a significant contribution to the BISTI effort.

Health Services Research and NLM's Training Programs

NLM is aware of the huge potential of informatics for handling the large datasets of health services research and public health

research. EP collaborated with Library Operations in the joint meeting of NLM with the Agency for Healthcare Research and Quality on Health Services Research in January 2000. To promote research training in health services research, EP distributed \$50,000 to each of the 10 NLM-supported training programs that requested such supplements.

Individual Fellowships

Individual informatics research fellowships are available to those who seek research training similar to offerings at the institutional training sites but at a site of their choosing. Individual applied informatics fellowships are also available to individuals who want to learn informatics techniques and technology for application in their current professional specialties. To encourage mid-career applicants, the applied fellowships permit stipends of up to \$58,000 per annum as substitute for salary lost during each training year.

Education of Health Sciences Librarians in Informatics

All NLM informatics training programs have been encouraged to develop and offer training within the curriculum suitable for those interested in health science libraries. NLM agrees to provide additional funding for any slots awarded to librarians. Response has been gratifying and is growing. Librarians are now in place at the University of Pittsburgh, Oregon Health Sciences University, and University of Missouri.

Research Support

Research support is provided through a variety of mechanisms, including individual research grants and contracts, cooperative agreements, research resource grants and others. NLM's research grants support both basic and applied projects involving the applications of computers and telecommunication technology to health-related issues in clinical medicine and in research.

Medical Informatics

Since inception of the grant program, the majority of NLM's research support in informatics has focused on the informatics of health care delivery with support both to applied projects (e.g., the electronic medical record, telemedicine) and related basic problems (e.g., natural language processing, data-mining, knowledge representation).

Biotechnology Informatics (Bioinformatics)

Clinical medicine remains an important area for NLM research grants, but NLM has been aware for a decade that the techniques of informatics are also indispensable tools for handling the complex data generated by research, most notably in molecular biology research and neuroscience, but also in clinically relevant areas such as outcomes research and public health issues. To facilitate this form of biomedical computing, EP has maintained a separate grant program (now called "bioinformatics"). NLM continues to provide research grants for informatics projects in this area, as well as training grants, and grants for support of research resources.

Databases

A related problem concerns the development and maintenance of electronic databases on which researchers increasingly rely, and for which no other source of support has yet been identified. NLM is an important, but not the only, NIH source of support through grants for such databases.

BISTI

NLM has been moderately successful in recent years in interesting other NIH Institutes in supporting informatics projects wholly or in partnership with NLM. However, the BISTI report markedly increased NIH interest in potential of computing for biomedical research. In FY2000, NLM together with a number of other Institutes began a continuing series of discussions about the various ways in which

NIH intends to address national needs for training and research in biomedical computing.

With participation by NLM and numerous other Institutes, NIH announced a battery of new programs responsive to BISTI in late FY2000 with the first awards to be made in FY 2001. The extent to which NLM will participate in funding or co-funding of BISTI projects remains to be seen, and depends in part on the nature of the projects that will be proposed by the applicants. NLM will undoubtedly continue support of BISTI-relevant training activities, as described above.

NLM and the Human Brain Project

NLM participates with 15 other NIH and federal organizations in the Human Brain Project, which is led by the NIMH and seeks innovative methods for discovering and managing increasingly complex information in the neurosciences. Each participant selects grants within the project for full or shared funding.

Other Grants

Publication Grant Program

The Publication Grant Program provides short-term financial support for selected not-for-profit, biomedical scientific publications. Studies prepared or published under this NLM program include critical reviews or research monographs in the history of medicine and life sciences; on special areas of biomedical research and practice; on medical informatics, health information science and biotechnology information; and in certain instances, secondary literature tools and scientifically significant symposia. Resources in recent years have been used principally for history of medicine projects. Standard print publication has been the most common format, but projects in electronic publishing, video, and other media have also been supported. The program has an informal self-imposed ceiling of \$50,000 on direct costs per grant per year.

Minority Support

NLM is participating in an NIH-wide Fellowship Program aimed at encouraging under-represented minorities in research careers. In FY2000 Dr Carlton R. Moore, an African-American physician at Mount Sinai Medical Center in New York, continues work on an applied medical informatics fellowship for work on a database of free medications for indigent patients. His mentor is Dr Joseph Kannry.

Internet Connection Grants were awarded to a broad variety of institutions serving African-American, American Indian, and native Hawaiian populations in inner cities and rural areas. Similarly, a number of Information Access Grants were awarded to organizations serving rural and inner city populations.

Pan-NIH Projects

NLM also participates in a number of other multi-institute projects including bioengineering, pharmacogenetics, imaging, and nanotechnology. In FY2000 NLM provided co-funding for a pharmacogenetics database under development at Stanford.

Conference Grants

Support for conferences and workshops is intended to help scientific communities identify research needs, share results, and prepare for productive new work.

Biomedical Ethics

Ethical issues in health care and research produce an enormous literature. This literature comes from law, medicine, public health, and government. The National Reference Center for Bioethics Literature at Georgetown University continues to offer invaluable resources and guidance for workers in this area. An NLM contract maintains the Center. A complementary contract from Library Operations supports an indexing activity that contributes to BIOETHICSLINE, one of NLM's online databases.

HPCC and Outreach

The outreach and the High Performance Computing and Communications initiatives of NLM are elements of the formal grant programs.

Special Projects

In addition to its standing grant programs, Extramural Programs Division participates in a number of special projects often involving cooperation with another NIH institute or other Federal agency. Some examples of such activities in FY2000 follow.

The Digital Libraries Initiative-Phase 2 (DLI-2)

This initiative explores innovative digital libraries research and applications. The program extends the previously sponsored "Research on Digital Libraries Initiative." The term "digital libraries" is used to denote the vast distributed collections of text and images available through the Internet. Much research and development will be needed before these new electronic libraries can be used easily and efficiently to obtain reliable information. DLI-2 is administered by the National Science Foundation and is jointly sponsored by the NSF, the Defense Advanced Research Projects Agency, the NLM, the Library of Congress, the National Aeronautics and Space Administration, the National Endowment for the Humanities, and others.

The project is interested in electronic information in a broad spectrum of fields in arts and science. Improving network-based information access for health care consumers is an important goal of the project for NLM, although all aspects of digital libraries as applied to health domains may compete for funding. NLM, as have the other sponsors, contributed funds to NSF, which will manage the project. NLM's commitment for FY2000 was \$1,000,000 as it had been in the previous year, and represents an arm of the HPCC initiative. Target for total project budget from all sources is \$50 million over 5 years.

NLM made available to interested applicants the Unified Medical Language System Knowledge Sources and the Visible

Human datasets. Applicants were also free to use resources of their own choosing. Although awards were not made to fill predetermined domain quotas, the review and awards process resulted in a gratifying number of projects with health themes, and several others whose informatics component concerned issues with considerable potential benefit for health concerns. All of the contributed funds are now being used to support the out years of grants awarded during the first round of competition. As of now, no surplus funds were available to support applicants who sent in proposals during round 2.

Informatics for the National Heart Attack Alert Program (Research Contracts)

This program receives approximately two thirds of its funding from NHLBI, and the remainder from NLM. The program offered a Phase 1 feasibility contract for up to \$100,000 for one year. Phase 2 called for implementation in a test population or a larger group over a period of several years. After the initial Phase 1 RFP in FY1998 which focused on "main-line" informatics and supported 14 investigators, a second Phase 1 RFP was published in FY1999 to obtain feasibility proposals using more innovative, high-risk, high-payoff technology. Five Phase 1 contracts for nine-month planning phases were awarded in this "high-tech" group. Technologies to be explored include wearable devices, portable computing devices, games, and wireless communications devices. In response to a Phase 2 RFP for the "main-line" Phase 1 projects, five Phase 2 contracts were awarded during late FY1999 and FY2000. A Phase 2 RFP for the "high-tech" projects was issued in late FY2000. Awards will be made in FY2001.

Miscellaneous Special Projects

NLM continues to transfer funds to other agencies to support projects of broad scope and utility for biomedical research. The agencies that received funds were National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institute on Deafness and Other Communication Disorders, National Institute of Mental Health, Agency for

Healthcare Research and Quality, and The National Science Foundation. NLM received co-funding for NLM grants from other organizations, including Department of the Army and NIMH.

SBIR/STTR (PHS 301)

All NIH research grant programs, including NLM's, by Congressional mandate allocate a fixed percentage of available funds every year to Small Business Innovation Research (SBIR) grants. These projects may involve a Phase I grant for product design, and a Phase II grant for testing and prototyping. NLM also participates in the other mandated fund allocation program, Small Business Technology Transfer, but generally it contributes its small allocation to other NIH institutes, as it did this year.

Grants Management Highlights

The Grants Management staff reviews NLM grant applications for administrative content and compliance with guidelines and directives; prepares and disseminates grant award documents; maintains official grant files for NLM; provides consultation and assistance to grantees on appropriate business management concepts; and advises NLM officials on grants management policy and procedures. The Grants Management staff, which consisted of three employees, issued a total of 159 awards for FY 00. New program areas included supplemental awards for the Biomedical Information Science and Technology Initiative (BISTI) and Health Services Research (HSR).

Review Committee Activities

NLM's initial review group, the Biomedical Library Review Committee (BLRC), evaluates grant applications for scientific merit. BLRC met three times in FY2000 and reviewed 108 applications. The Committee (see Appendix 5 for roster of members) operates as a "flexible" review group; i.e., it is composed of three standing subcommittees: 8 members on the Medical Library Resource Subcommittee, 9 members on the Medical Informatics

Subcommittee; and 4 members on the Biomedical Information Subcommittee. The subcommittees consider research applications in medical library projects, medical informatics, and biotechnology information respectively.

A second peer review of applications is performed by the Board of Regents, which also meets three times a year, approximately three months after the Biomedical Library Review Committee. (A roster of members is in Appendix 2.) One of the Board's subcommittees, the Extramural Programs Subcommittee, meets the day before the full Board for the review of "special" grant applications. Examples include applications for which the recommended amount of financial support is larger than some predetermined amount; when at least two members of the scientific merit review group dissented from the majority; when a policy issue is identified; and when an application is from a foreign institution. The Extramural Programs Subcommittee makes recommendations to the full Board, which votes on the applications.

Thirteen Special Emphasis Panels (SEPs) were also coordinated which reviewed 69 applications. Such panels are convened on a one-time ad hoc basis to review applications for which the regularly constituted review groups lack expertise. Use of SEPs by EP increased significantly in FY2000 because of new regulations requiring EP to supervise SEP for review of contracts as well as grants.

Personnel Activities

EP has three Program Officers, one each for the three areas of Library Resource, Informatics, and Publications. These staff members work with grant applicants during all phases of the application and review process, and subsequently monitor the work done on the awarded grants. They are an important "interface" of NLM with the academic community. After decades of service, the Program Officers for Library Resource and for Informatics retired in FY 2000. They are missed as colleagues, and the loss of their expertise will take much time to restore.

A new Administrative Officer was hired in FY2000 to replace the previous incumbent

who retired for health reasons. EP also hired a Program Assistant, a Committee Management Assistant, and a Grants Technical Assistant (Review) to aid a variety of management functions with emphasis on the demands of IMPAC II, NIH's new database system. In addition EP recruited for a Program Assistant, and a Committee Management Assistant for FY2000. Mrs. Frances Howard, long-time special assistant to the Director, EP, retired in FY2000.

Summary

EP's grant activities in FY2000 were in conformity with previous years with the exception of the significant new emphasis on

biomedical computing as stimulated by BISTI, and a relatively modest but seminal specific expenditure in support of informatics training for health services research. NLM's extramural grant division, like similar divisions elsewhere at NIH, cannot fund all applications of good quality, but the grants which can be made are furthering NLM goals in most key areas. The situation was particularly acute in FY2000 because a larger than average demand on the available budget came from the budget requirements of existing grants. A number of excellent grants, both in informatics research and in library resource grants were held over to FY2001 in the hope that the new budget will permit funding.

Table 10

Extramural Grant and Contract Program (Dollars in thousands)

Category	FY 1998		FY 1999		FY 2000	
	No.	\$	No.	\$	No.	\$
Resource projects	39	5,897	48	6,339	28	5,604
IAIMS	9	3,690	11	3,611	11	2,824
Access	3	198	8	1,025	9	696
Systems	10	1,352	9	1,025	13	1,878
Connections	17	657	20	678	5	206
Research	72	15,887	87	53,635	61	24,471
Medical Informatics Projects	38	8,384	43	12,000	45	12,567
Medical Informatics Resource
Biotechnology	16	4,730	6	2,972	7	2,973
Cooperative Agreements
Career Awards	17	1,773	11	1,199	7	7,781
Library Science
Digital Library	1	1,000	27	37,464	2	1,150
Training	23	5,519	21	1,239	23	8,217
Institutional	12	4,993	12	758	11	7,419
Fellowship	11	526	9	481	13	798
Publications	8	269	5	222	6	267
Bioethics	1	513	1	528	1	530
SBIR/STTR	5	508	5	528	5	424
Regional Medical Library	8	6,710	8	8,706	8	8,185
NIH Tap		830			1,030
Totals:	156	\$36,133	167	\$71,197	142	\$48,728

OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

Simon Y. Liu, Ph.D.
Director

The Office of Computer and Communications Systems (OCCS) provides efficient, cost-effective computing and networking services, application development, technical advice, and collaboration in informational sciences in support of the research and management programs offered through the NLM.

OCCS develops and provides the NLM backbone computer networking facilities, and supports, guides, and assists other NLM components in local area networking. The Division provides professional programming services and computational and data processing facilities to meet NLM program needs; operates and maintains the NLM Computer Center; designs and develops software; and provides extensive customer support, training courses and seminars, and documentation for computer and network users.

OCCS helps to coordinate, integrate, and standardize the vast array of computer services available throughout all of the organizations comprising NLM. The Division also serves as a technological resource for other parts of the NLM and for other Federal organizations with biomedical, statistical, and administrative computing needs. The Division promotes the application of High Performance Computing and Communication to biomedical problems, including image processing and information security.

Overview

This year, OCCS was reorganized to streamline operations and enhance customer support. Roy Standing, Acting Director through January 2000, planned and initiated the reorganization; while Joe Hutchins, Acting

Director from February to May 2000, carried out and executed the reorganization plans. Several new managers were selected during the year. Simon Y. Liu, Ph.D., was selected as the new Director of OCCS. Michael Bumbray, Kathy Cooper and Garry Fox became the Section Heads of Facilities Management Section, Desktop Services Section, and Systems Services Section, respectively. The new OCCS organization structure has the Applications Branch and the Systems Technology Branch, which together have a total of six sections.

The Applications Branch is responsible for providing technical support for many of the NLM's national and international programs. The branch provides requirements definition services, recommends system architectures and implementation strategies, designs and develops computer solutions including hardware and software, plans and implements database management solutions, maintains and enhances existing systems, provides consulting services to NLM related to information technology, and evaluates emerging technologies. The two sections of the Applications Branch and their key activities are the:

- Information Management Section, which provides support for many NLM databases and services. The section played a major role in the NLM's System Reinvention project and will continue to provide database and application support in the new environment. IMS is responsible for administrating databases and related applications, providing technical solutions, and maintaining systems in order to improve system features and capabilities.
- Software Support Section, which provides support for systems development and consulting services both within NLM and for other organizations. This Section is responsible for analyzing user requirements, evaluating and selecting appropriate computing solutions, creating data and workflow models, and software development resulting in operational systems.

The Systems Technology Branch provides technical support for information

technology systems that are provided by OCCS. The Branch is responsible for security, availability, and reliability of supported systems. It maintains a Help Desk that offers staff and contractors a single point of contact for reporting problems and making requests for services. The sections of the Systems Technology Branch and their key activities are the:

- Systems Services Section, which is responsible for administering, managing, and efficiently utilizing NLM/OCCS computer equipment, and its operating systems and associated software. The Section supports the IBM mainframe hardware and systems software and Unix-based systems. The Section also implements new or improved operating systems and associated software, and selects and acquires computing and storage equipment and features.
- Facilities Management Section, which provides 24-hour, 7 day-a-week, 365 days a year, staffing, system monitoring, immediate response and system support services to users within OCCS and other NLM organizations. In addition, this Section provides steady, filtered, independent electrical power and assigned raised floor space for its users' computer hardware. The Section is responsible for functions and duties that include: batch processing, batch scheduling, system monitoring, problem and system analysis, problem and system reporting and system intervention.
- Network Engineering Section, which deploys, administers, and maintains local and wide area networks. This Section manages network services for NLM, including all network-related issues including future plans, evaluation of equipment and management of procurement of network systems.
- Desktop Services Section, which provides NLM-wide software user assistance, administrative workstation support, and workstation hardware platform integration, upgrades, and configuration coordination. This Section also ensures that all NLM Divisions receive effective and timely advice, guidance, and technical support for

their personal computers and workstations. The Section manages the delivery of desktop hardware and software services throughout the NLM.

As last year, OCCS has been maintaining legacy systems while developing and transitioning to new systems that incorporate new technologies and re-engineered approaches. OCCS strives to continue to improve customer service by encouraging staff to be proactive and do more for its internal and external customers. Major OCCS milestones and successes this year included:

Infrastructure Improvements

- The elimination of the Value Added Network (VAN) connections. VAN services are no longer required since the new version of DOCLINE is now accessible via TCP/IP, unlike the previous mainframe-based version of DOCLINE.
- The completion of the upgrade of the NLM LAN infrastructure from 10BaseT Ethernet to 100BaseT Ethernet, using 100Mbps Cisco Switches.
- A consolidated NLM computer facility is now available with more usable floor space and electrical power meeting the specifications of various Unix and client/server hardware. This year mainframe peripherals, old dormant mainframe power and cabling, and the 18 inch raised floor were removed.

Customer Support Enhancements

- OCCS provides continuous (24 hour and 7 day) monitoring, reporting and human intervention for LHCBC, NCBI, and SIS systems that have been moved to the new NLM computer facility.
- No magnetic tapes will be used for year-end processing of MEDLINE. Since web-based retrieval and DLT tapes are to be used, over \$100,000 in tapes and processing has been saved.
- Smooth Year 2000 transition. The extensive testing and preparation for Y2K paid off. OCCS had extra staff in the computer room

around the clock to ensure the smooth transition.

Reinvention Systems Advances

- A new web-based version of DOCLINE was introduced that integrates three previously separate subsystems and includes interfaces with PubMed, Voyager and IGM. First year savings of \$200,000 are anticipated.
- An initial DCMS operating capability was deployed successfully this year. Eventually, DCMS will replace AIMS and support all journal articles that are indexed with MeSH headings. From October 1, 2000, onward, all data that was previously keyboarded or scanned will be received in XML.
- OCCS was heavily involved in the conversion of monographic data from several ELHILL databases into the Voyager database. This is a key step in the eventual replacement of this legacy system, which has been maintained for the last 25 years by OCCS.

Consumer Health Information Applications

- The very popular MEDLINEplus had several new releases this year. New features include a medical encyclopedia, extensive drug information, and a new design that features a more attractive and compact home page with five major content areas.

New Administrative Support Systems Applications

- OAMS Inventory system, an NLM internal system that allows users to order office supplies online and automates the inventory management process, was introduced this year.

The following describes, in detail, OCCS's many accomplishments achieved in each major functional area for FY2000.

Customer Services

As part of the reorganization of OCCS, all OCCS support is now provided and tracked through a single point of contact, the IT Services Center (ITSC). This focus builds on the IT

Services Center concept that OCCS introduced last year. The ITSC now provides NLM users with a single, reliable point of contact for all supported products and services within OCCS. To assist management in tracking operations, a daily IT Services Center status report is e-mailed to a wide range of users and managers. It provides a summary of the status of all OCCS systems and any status changes that have occurred. The report also provides summary statistics on requests for services.

This year, OCCS exceeded its Service Level Agreements goals and reported no unscheduled downtime of the mainframe. Other customer services initiatives encouraged OCCS staff to become more proactive in providing customer service, such as providing more first-call problem resolution of trouble calls/requests placed to the Services Center.

Y2K Readiness

The extensive testing and preparation for Y2K paid off. OCCS led the coordination of the NLM's desktop Y2K identification and remediation activities. Three OCCS employees leading the endeavor received the NIH Director's Award for their efforts. The NLM-wide Y2K project team identified essential software and hardware to prepare for Y2K preparation efforts, which included the removal and surplus of non-remediable hardware. Over 250 software titles were researched for Y2K compliance. Over 1,500 PCs and their installed software were audited for failing software versions. The team successfully remediated the hardware and software.

Contingency plans were developed by OCCS for three mission critical systems, namely TOXNET, PubMed, and MEDLARS Database Updating. These three mission critical systems passed all the Y2K compliancy checkpoints. All other NLM systems were up and running as expected January 1, 2000.

In addition to the testing and IV&V that were performed, both Day One and Leap Day plans were prepared and executed. Various backup scenarios were in place, including running one Rescue System one week "backwards" in time, to be prepared if a Y2K meltdown happened.

There were no significant problems with communications, mainframe, LAN, Unix, or desktop systems as a result of the calendar change from 1999 to 2000. All major applications and software executed without incident. Many OCCS staff were on site or on call but, thankfully, had little to do other than monitoring and reporting. Extra staff were on hand in the computer room around the clock just in case. This effort led to the smooth Y2K transition at NLM.

Desktop Support

Service Level Agreement goals were exceeded this year, and OCCS continued to expand its use of standards and automated tools to provide improved services. The IT Services Center is open Monday-Friday from 7:30 AM to 5:00 PM and is staffed by a combination of Government employees and contractors. During FY2000, the IT Services Center handled approximately 6,000 trouble tickets, 2,461 telephone calls, 418 e-mails, and 153 walk-ins. The average resolution time was 2.3 hours for High priority tickets, 1.75 days for Medium priority tickets, and 7 days for Low priority tickets. This is less than half the required resolution times stated in the OCCS Service Level Agreements, which are 8 hours for high priority, 2.5 days for medium priority, and usually a week, for low priority.

Workstation Installation and Software Training

A key responsibility of the Desktop Services Section is the configuration and installation of new and previously owned PCs and peripherals. Approximately 450 workstations were installed for NLM customers this year. DSS also oversees in-house PC software training for NLM staff. A total of 516 students attended 58 classes on Microsoft Office 97 applications, GroupWise and the NT Operating System. A training pilot with the National Institute of General Medical Sciences (NIGMS) was successfully completed this year. A permanent agreement for OCCS to provide training to NIGMS was signed for FY2001. In addition to classes, over 75 individual one-on-one visits were made to staff who needed help in

a specific software application area but did not feel a full class was warranted. New training offerings are planned for next year, which will include a class on the differences between Office 97 and Office 2000 and a "training buffet" where individual topics in software applications can be selected to tailor the needs of a specific section.

Application Distribution and Desktop Management

The Novell Application Launcher (NAL), an automated PC software deployment tool, continues to be used to distribute applications, patches, and updates to hundreds of NT desktops in NLM. This automatic deployment of thousands of updates and applications has saved many FTE hours of work. Since the NAL can deploy only to workstations that are part of the Novell environment, other distribution methods are being explored for other platforms.

For the first time, a complete PC hardware and software inventory is being collected on all Novell-attached systems. The inventory was collected using the second version of ZenWorks, which was released in 2000. Besides aiding in PC asset tracking, such inventory information will enable more efficient maintenance of desktop software. Also, when responding to calls, the IT Services Center will know what software is on the workstation of the user requesting service. These tools have many more features and are an integral part of the Novell Network Directory System (NDS) family of products.

Extensive testing of the Windows 2000 (W2K) desktop operating system was conducted this year. W2K was found to be a generally good desktop operating system. However, testing revealed that several key applications, including the latest release of Voyager, are not supported under Windows 2000 at this time. Therefore a limited deployment was recommended and approved. To date, Windows 2000 has been deployed on 5% of the PCs in NLM. W2K provides more stability and introduces W2K Server Active Directory, a directory which stores information about network components. The testing of the Microsoft Office 2000 suite is

under way, and implementation will begin in FY2001.

Workstation Policies, Standards and Cost Savings

In May 2000, NIH instituted a Sanitization Policy for hard disks. The policy states that before any NIH-owned or managed hard disk or system containing a hard disk is transferred, surplus, or donated, it first must be sanitized either by reformatting the hard drive in a secure manner or by using an approved wipeout utility. The DSS provides this service for NLM PCs. Approximately 30 PCs have been sanitized during this fiscal year. An estimated \$85,000 in direct and administrative costs were saved this year by using consolidated acquisitions to purchase PCs based on standard specifications produced by the LO/OCCS Personal Computer Advisory (PCA). Approximately 200 PCs were acquired this year following the PCA methodology.

In the future, DSS intends to provide more services to customers to meet their computing needs. Key services under exploration include, computer maintenance, problem resolution of trouble calls, providing statistics from the Help Desk database.

Network Support

During FY2000, the Network Engineering Section (NES) LAN and Communication systems area continued to provide reliable LAN and Internet communications services, meet the communication needs of newly reinvented systems, explore new technologies and plan for NLM's continued growth in communications. Looking forward, NES is taking further steps to increase capabilities of networks and of storage by providing for better performance, more redundancy, and enhanced backup and storage.

Network Infrastructure Upgrade

During FY2000, the performance of the NLM LAN infrastructure improved significantly as the upgrade from 10BaseT Ethernet to 100BaseT Ethernet was completed. Over 60

10Mbps Cabletron hubs were replaced with 100Mbps Cisco Switches. Now, over 1600 connections supported by NES are switched 100Mbps connections. In addition to connections among the core devices, this upgrade included printers, desktops and servers connected in the computer room and communications closets.

Network Management

The critical ability to monitor the current status of NLM network components was enhanced during 2000. For the past several years, SunNet Manager has been used to monitor the major network routing nodes. Since it and other tools did not provide a full picture of the network and its devices, HP OpenView, a fully featured management platform, was selected and is now used by staff to manage and monitor a wide range of hardware and software. In particular, the Network Node Manager application of HP OpenView is being used to monitor and troubleshoot the OCCS supported parts of the NLM network. Not only are network communications devices such as routers and switches and various Unix, Novell, and NT servers monitored by OpenView, but operating systems, Oracle databases and various NLM applications are also being monitored using OpenView.

Other monitoring packages for specific systems were also implemented this year, including MailCheck and MailCentral to monitor the GroupWise e-mail system, Compaq Insight Manager and Dell OpenManage for Dell and Compaq servers; and DS Expert and DS Analyze for Novell NDS.

Novell NetWare Operating System

The majority of Novell Netware systems were upgraded to Netware 5 during 1999 and 2000. Only three legacy 3.12 systems and two Netware 4.11 systems are still running. All of these systems will be retired in 2001. A major project that will increase storage capability on Netware and provide redundancy for critical services is the Storage Area Network (SAN) project, scheduled to be implemented in FY2001. The SAN hardware provides the

centralized data storage and the NetWare software enables a secondary server to provide a service even if the primary server fails. This SAN will also increase the amount of file storage available to users to over 200 gigabytes. This additional storage space will permit centralized backup of user files. A future use of the SAN technology for OCCS is to explore using it as a shared storage device across multiple platforms and operating systems. The SAN can support NetWare, NT and Unix files concurrently and centrally backup those file types. This will be tested and possible implementations reviewed for viability.

GroupWise E-mail System

The GroupWise system grew to over 900 users and the 5.5 enhanced GW client was distributed this year. Among other new features, the 5.5 client supports digital signatures and LDAP lookup. The web access was also upgraded to the Enhancement pack that included new cryptographic services (SSL) on the GW GWIA server. GroupWise has been very stable and dependable, and the addition of both MailCheck and MailCentral has enabled LAN Support to detect problems early. Recent releases have provided better administrative options that make GroupWise easier to support. In FY2001 GroupWise version 6.0, called Bullet Proof, will be released.

Microsoft NT

The Microsoft NT Domain continues to grow in importance in the OCCS environment. NT servers are used for testing and running a number of Reinvention applications. IIS is used to host the MailCheck and Insight Compaq Manager. RELAIS is NT based as are services such as DHCP, Web browsing, and NT Authentication. The NT Domain will undergo further changes as the Windows 2000 server and the Active Directory are tested. Active Directory, which is now running in the OCCS lab, requires additional testing for requirements such as DDNS (Dynamic DNS) that are not easily implemented.

Extramural Programs Support

The IT functions of the NLM Extramural Programs (EP) Office continued to be supported in EP's off-campus location in the Rockledge I building in Bethesda. On-site technical support is provided for the PC, network, and IMPAC II systems.

Internet Connectivity

Production Internet connectivity services continued to be provided through a contract with Genuity (formerly GTE/BBN Planet). This NLM contract provides T3 (45Mbps) connectivity to the GTE/BBN Planet network node in Washington DC via a SONET ring. It also provides a T3 link for CIT/NIH to the node in Vienna, Va. NLM and NIH collaborate in using these links to backup each other's Internet connectivity. PSC and NCI links to the Internet are also provided through this contract. A new 4-year contract was awarded for Internet connectivity services for NLM, NIH, and PSC/HHS. NLM plans to upgrade the T3 link to an OC3 (155 Mbps max) in FY2001.

Wide Area Networks

OCCS terminated the Value Added Network (VAN) connections provided by Tymnet (MCI) in August 2000. VAN services are no longer required since the new version of DOCLINE is fully accessible via TCP/IP. Some NLM users had required VAN access since the previous mainframe-based version of DOCLINE was not fully functional with TCP/IP access. The termination of VAN access is a milestone in technological advancement. NLM has had to provide some form of VAN access—either through Tymnet, Telenet, CompuServe, or FTS2000—ever since its services have been available online. VANs have been needed to support MEDLARS since 1970. Now virtually all communication is accomplished through the Internet.

Remote Access

Network support continues to provide 56K dial-in access, Remote Access Server to

NT, and ISDN access for a wide range of NLM users. One technical challenge overcome this year was the replacement of the mainframe AIMS system with DCMS. Many of the support contractors, however, still have dumb terminals or low-end PCs. For continuity of support with the new system, a high-speed communications solution for these remote Indexers had to be found. A Windows-based terminal solution was tested. While this solution provided a fault tolerant workstation for remote users with centralized support for software and configuration, the new application was found to be somewhat slow on connections less than 56K.

Consequently, OCCS recommended DSL as the most effective solution for high-speed access to DCMS. Because DSL is not yet available at many users' locations, OCCS intends to provide a dial-in service for users who do not qualify for DSL. DSL does require the use of additional security software called a VPN (Virtual Private Network) client, which has been installed.

In addition to supporting the indexing system, a terminal server setup has been tested and found to be a good fit for flexi-place workers. A separate file server is being purchased by Library Operations for that purpose. In 2001, NLM dial-in services and DSL services are planned to be expanded to replace the existing remote access offerings.

System Support

FY2000 was another transition year for OCCS system support. Staff continue to be responsible for maintaining legacy systems until they are phased out while simultaneously deploying and maintaining new client/server applications. In addition, staff are planning for the future by constructing the infrastructure needed for additional Unix systems. More than 60 Unix systems already are supported. As part of the OCCS reorganization this year, mainframe and Unix systems support teams were merged into one reporting unit to ease this transition. The main system support activities for FY 2000 included:

- Installation, maintenance, and support (IMS) of NIS, NFS, DNS, and Web services;
- Unix O/S IMS for more than 60 systems;

- Hardware IMS for more than 60 systems;
- Monitoring, Performance, Analysis and Tuning for more than 60 systems;
- Oracle database IMS for 23 applications;
- Security and Account administration for more than 60 systems;
- Reading Room support for several dozen workstations; and
- Legacy O/S, program product, and application support.

To provide this increasing range of system support, OCCS is relying increasingly on Configuration Management. The new Unix systems and applications are complex. As these systems move from the testing/development phase towards production status, it has become apparent that configuration management together with supporting tools must be used more in the OCCS production environment. This is now beginning. For example, use of additional features of the CA/Scheduler for Operations has reduced the support staff requirements for the mainframe. This has freed up facilities management staff to monitor Unix machines. Even though much work remains to be done, significant progress is being made in the areas of change control, software promotion, logging, maintenance, and documentation.

As part of transition from the legacy systems, key milestones this year were:

- DOCLINE, one of NLM's oldest systems, is now off the mainframe.
- The stabilizing and downsizing of the IBM 9672/R52 mainframe continued this year as various applications and services, including the Model204 and DOCLINE, were deactivated. While the support necessary to keep the mainframe running was performed, few enhancements other than mandatory Y2K preparations were required this year. Since the mainframe is projected to be phased out by the end of FY 2001, extensive patching of mainframe software, e.g., new versions of mainframe TCP/IP, are not needed.
- All SAA Gateway users and printers within NLM were converted this year to a TN3270 protocol using TCP/IP as an Internet connection to CIT. Since 1994, the SAA

gateways has provided terminal emulation access to NLM and NIH/CIT mainframes for a variety of applications such as ELHILL, CICS, TSO, and DelPro. The SAA Gateway, however, was not Y2K compliant and had limited capacity. As OCCS mainframe services continue to be reinvented, such NLM mainframe Internet access has decreased by approximately 50%.

Supporting the systems of the future, system support staff this year have:

- Deployed eight Sun E250/400 workgroup servers and moved Unix applications to them;
- Assisted in converting RELAIS Unix to the latest Sun operating system;
- Served as Unix system administrators;
- Deployed the HP OpenView software monitor (for IT Operations and plugins) on all our databases and production servers as well as on client workstations;
- Began maintenance for new Unix software packages such as ColdFusion; and
- Supported ColdFusion load testing.

The personnel transitioning that began in the system support area last year continued as it did throughout OCCS. Some staff members transitioned to network engineering support after attending classes. Other legacy support staff, although still required to provide legacy systems support, transitioned to either management or client/server support roles. Various classes and on-the-job training were needed to support this transition. Not unexpectedly, staff members found it difficult but possible to perform these dual concurrent roles of “supporting the old” and “training for and supporting the new.” One effective learning technique used as part of the transitioning process was to pick a process or task and document it to reach the point of understanding it.

System Security

Due to an aggressive approach to security and intrusion detection, NLM’s overall security measures have been exceptional.

Improved security software, including superior intrusion software, was installed in 1999 and relied on extensively in FY2000. An improved, highly performing firewall is now in place. Several firewall packages had been tested and one was chosen, leading to better monitoring and firewalls.

NLM network security includes both firewall and intrusion detection technologies. NLM’s intrusion detection system monitors all inbound traffic from the Internet for suspicious activity and generates alarms when a set of circumstances is met. NLM’s firewall filters out traffic based on certain rules specified for known security threats. Thorough virus protection at NLM relies on tools for each of three levels: at the incoming T3 line, in the Post Office, and at the desktop.

The McAfee Virus scanning software is currently implemented on all supported PC desktop systems. All GroupWise e-mail is scanned for viruses as e-mail is received from the Internet. This year, the Guinevere virus scanning for GroupWise was implemented and a new post office was created for the Remote Indexing Contractors. Due to the internal design of GroupWise and the virus scanning of all incoming e-mail, NLM maintained e-mail services during the entire worldwide ILOVEYOU virus problem, unlike many other sites.

NLM has been very successful in intrusion prevention. Security staff has successfully uncovered several hackers. Some have led to joint investigations with the FBI, a few of which led to convictions. One hacker detected had successfully attacked over 200 corporate and government sites without detection. It was NLM who caught him. Besides the security tools noted above, SATAN-like scans are performed regularly as part of monitoring systems. Scanning packages like SATAN (Systems Administrators’ Tool for Analyzing Networks) are relied upon to scan Unix systems and TCP/IP networks to counteract the much more powerful scanning tools already in the hands of hackers. NLM is well represented in security circles at NIH and Government-wide.

Computer Facilities

NLM systems are supported in a safe, secure environment in NLM's Computer Facility, which is open 24-hours-a-day, 7 day-a-week for 365 days a year. OCCS staff provide system monitoring, immediate response and system support services to users within and outside OCCS and other NLM organizations. A big step this year was completion of the upgrade to the NLM computer room, which provides better monitoring, more space, and improved power support.

A consolidated NLM computer facility is now available with more usable floor space supported by electrical power meeting the specifications of Unix and other client/server hardware. New Unix and client/server systems have been moved to the computer facility from LHNCBC, NCBI and SIS. These computers previously had been located throughout the building. Key aspects of the upgrade to the Computer Facility were that:

- Approximately 1500 square feet of useable floor space became available for other use when mainframe peripheral hardware and furniture were surplus.
- Tiles for the 72 x 90 square foot, 18 inch raised floor were replaced throughout the entire computer facility, and the floor was leveled out.
- Old dormant mainframe power and cabling also were removed. New electrical power was added to the computer facility to meet the specifications of Unix and client/server hardware. As more Unix servers are brought in, additional power lines, which have different phases, amperage and connectors, are being installed. The computer facility continues to be supported by Uninterrupted Power Supply.
- OCCS also has been reorganizing the placement of equipment to free up power panels. Staff have identified and updated documentation on each computer systems power source, and re-labeled each circuit within each electrical power panel. Moving the electrical circuits within the computer facility will provide each NLM organization with a separate power source. This will

enable them to install new systems more quickly and to isolate problems much better. The reorganization of power sources is almost to the point where each organization has its own panel.

- All connections were documented and cable management was installed in the communication racks as part of the upgrade. Most of this work was done after-hours so as not to interfere with end-users' ability to work. Cable management has proven to make a significant difference in identifying and troubleshooting cable problems.
- A test desktop and network lab was established in the rear of the computer facility. New systems such as Windows 2000, AD, NW 5.1, and GroupWise 6.0 or new technologies can be tested prior to implementation. The intention is to permit other NLM Organizations, as well as OCCS, to use the lab.

OCCS is moving steadily to a complete facility management approach. Other NLM divisions besides LO have been provided with floor space, power, system monitoring, system interaction and system reporting on an around-the-clock basis. Therefore, OCCS staff serve as a first line of contact when reporting major or critical service interruptions during off (non-traditional) business hours. Both OCCS staff and divisional staff monitor during the day. As OCCS staff learn more about these systems, they can increasingly intervene successfully rather than merely alert the other divisions to problems. The computer facility also has 24 by 7 physical security.

Newly installed monitoring tools and alarms are used to monitor these applications in the NLM Computer Facility. Support for NLM security has increased. Audible alarms have been programmed to signal intrusions so that intruders can be trapped and damage to applications prevented. Key accomplishments in systems monitoring include:

- Successfully automating all mainframe nightly batch-processing activities. This allows facilities management staff to concentrate on Unix and client/server processing activities. CBOMS, an automated Job Scheduler, is used to run automated

scripts instead of keying in information each time.

- Increasingly moving beyond passive monitoring. OCCS staff are learning how to deal with a wider range of problems immediately instead of just passing them on.
- Taking over a number of monitoring/checking activities previously performed by network support, such as monitoring the status of Novell and NT servers and the GroupWise mail flow. Facilities staff now report any issues or problems to LAN Support both during and after working hours. This has proven to be an excellent addition to facilities support.
- Handling day-to-day RELAIS application processing activities, including monitoring the system throughout its 22-hour daily processing schedule. Using the restart capability of RELAIS, facilities staff can completely recycle it during off-hours. Since OCCS monitors, reports and intervenes when necessary during the day-to-day application processing of RELAIS, LO's technical support personnel can concentrate on other endeavors.
- Successfully processing and shipping year-end processing for MEDLINE licensees. This is the last year that magnetic tapes will be used for year-end processing. In the future, Internet downloads and DLT tapes will be used.

Standard processing each weekend by facilities staff includes complete system pack and database off-site backups. The Tuesday immediately following each weekend, the backup tapes were shipped to a secured class A vault for storage. No unscheduled downtime for

mainframe systems occurred this year. Since the installation of UPS, unscheduled downtime has become very rare.

The mainframe-oriented staff are also undergoing career transitioning. One staff member became a certified Unix System Professional. Half of the current staff is expected to be certified in FY2001. As a result, they have been transitioning from day-to-day mainframe activities to Unix and client/server activities, supporting systems such as RELAIS and Internet systems.

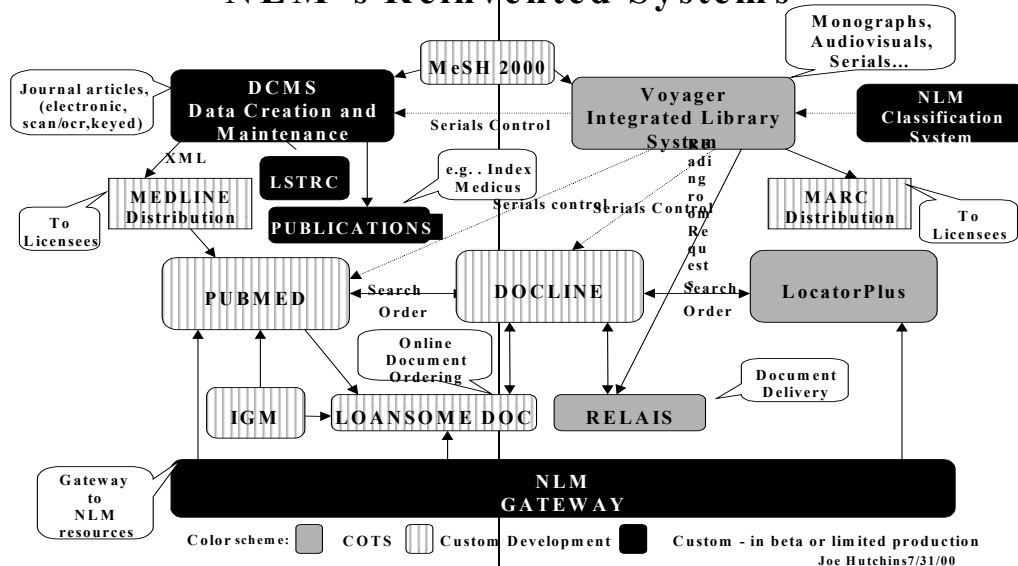
Reinvention Systems

A special NLM Systems Reinvention Milestones Reception was held at NLM on August 9, 2000, to celebrate three NLM System Reinvention landmark achievements:

- A web-based client/server MeSH2000 environment, which is more compatible with the UMLS Metathesaurus;
- The new web-based DOCLINE, which includes SERHOLD and DOCUSER; and
- The first significant phase of the new operational web-based Indexing Data Creation and Maintenance System (DCMS), which will replace the AIMS legacy system.

These are three of the many reinvention systems that OCCS is leading the way in designing, developing and implementing. The Reinvention systems are integrated as shown on the Overview graphic "NLM's Reinvented Systems," which shows the linkages between, and the development stage of, 15 systems.

NLM's Reinvented Systems



It is important to note that OCCS is reinventing many core NLM systems without any interruption to production systems. Other NLM divisions are developing some of these systems, e.g., PubMed and the NLM Gateway. In addition to the systems, the Reinvention project also is changing the media of NLM products and how NLM operates:

- Several NLM products that were previously paper or magnetic tapes (for MEDLINE licensees) now will be available as web downloads.
- By creating and maintaining all MEDLINE records, OCCS provides the key data that other major NLM reinvention systems such as PubMed rely upon.
- OCCS also must transition staff as legacy systems are transitioned.

During FY2000, progress was made in all of the systems shown on the chart. The specific OCCS accomplishments for these systems are described below.

Integrated Library System (ILS)

Much progress was made in implementation of Voyager, NLM's Integrated Library System (ILS), this year. The implementation of Voyager, a commercial package developed by Endeavor Information Systems, Inc., was a key milestone in NLM's System Reinvention efforts. The major ILS accomplishments this year were:

- Upgrades to Voyager, including a Y2K fix;
- Development of a custom-designed Discharge System to support re-shelving and additional check-outs during the day, which is a capability not provided by Voyager; and
- Several major data conversion efforts, including conversion of ELHILL data into Voyager and merging NCBI Journal Data into Voyager.

Major ILS Upgrades

OCCS performed a major upgrade of the Voyager system this year, learning several major lessons in the process. The upgrade to Voyager to Release 99.1, which included some Y2K fixes, was completed in the second quarter. The upgrade process, however, was complicated by the many custom interfaces that NLM has added to meet the full range of NLM requirements. Through close coordination with Endeavor, OCCS management and the ILS team worked together successfully to solve the upgrade problems quickly and effectively.

The complexities of completing and testing the Voyager Release 99.1 upgrade, however, indicated that an approved upgrade approach is required for future releases. To this end, OCCS established a Voyager Test System to test future releases of Endeavor as well as releases of the other commercial software packages that interface with the ILS. The planned test environment will duplicate the production system, thus supporting testing of

new releases and upgrades in a simulated production environment. The next upgrade version for Voyager system will be Gold 2000, which is expected in the first quarter of 2001.

ILS Discharge System

OCCS developed a custom Discharge System for the Collection Access Section of LO this year. The standard Voyager ILS does not support efficient re-shelving and re-checking out of library materials during the day. Previously, re-shelving could only be performed when the reading room was closed. The new OCCS-developed Discharge System, which was deployed March 2000, allows re-shelving at any time during the day. This distributes the workload and makes more effective use of resources. The production version of the Discharge System has been functioning well.

ELHILL Databases converted into the Voyager ILS

OCCS was heavily involved in the conversion of monographic data from several ELHILL databases into the Voyager database this year. This conversion is a key step in the eventual replacement of ELHILL. This major legacy system, which has been maintained for the last 25 years by OCCS, has a number of databases containing information about a wide range of biomedical subjects. The conversion effort is quite complex because unique conversion specifications, processes and programs are required for each of the ELHILL databases. Therefore, conversion of each database requires significant analysis by LO, followed by additional programming by OCCS to export monographic data from ELHILL databases into Voyager.

This year the joint team of LO and OCCS staff members successfully converted the HSTAR, HISTLINE, SPACELINE, and BIOETHICS databases, which contain, respectively, Health Start information, History Citation Data, Space (NASA) Data, and Ethics Data. This data, formerly available through ELHILL, is now available online through LOCATORplus.

This year, OCCS development team spent significant effort on the parsing and conversion algorithms that are at the heart of the conversion programs. In addition, OCCS staff members refined the process to support the new DOCLINE to Voyager data exchange. The conversion of the POPLINE database is scheduled to be completed during the first quarter of 2001.

Merger of Journal Data from NCBI databases into the Voyager ILS

Merging journal data from the NCBI databases into the Voyager ILS was another key milestone this year. Working closely with NCBI, OCCS produced data load programs and reviewed databases and reports. The merge programs maintained title aliases and two forms of key data fields, one normalized for comparisons and one in the original form for copying into ILS. OCCS tested the merge process thoroughly by generating NCBI data and testing merges for ISSN and ISO title abbreviation and special case handling. This will relieve NCBI of the burden of maintaining a separate Journal Authority Database for PubMed. Data will be created and maintained by LO and then exported to NCBI. In September 2000, OCCS performed the final steps, creating a new MARC file containing the records with merged data. LO has verified the merged data, which has been merged into the ILS.

MeSH 2000

The new web-based MeSH 2000 went live in November 1999. This was one of the major steps in the NLM Systems Reinvention Project. The previous MeSH legacy system was developed more than a decade ago. MeSH is NLM's controlled vocabulary thesaurus. As part of the reinvention project, the underlying data structure of MeSH was altered to afford a concept-based representation that is more compatible with the UMLS Metathesaurus. Together with the new DCMS, this system will simplify the annual maintenance of MEDLINE records. Since the MeSH vocabulary changes slightly each year, maintaining older records is a concern.

The MeSH2000 client/server architecture was implemented using current versions of Sun's Java DK and JFC. The client program runs on platforms that support this software, which include Windows 95, Windows NT, and Sun Solaris. The client exchanges data with the Oracle database server using stored procedures and SQL queries invoked through JDBC, the Java standard for database communication. The MeSH2000 server runs on a Sun Sparc Ultra-2 under a SunOS version of Unix with an Oracle DBMS. The server portion of the software is implemented using Oracle data structures written in PL/SQL.

New DOCLINE

On July 17, 2000, NLM's new web-based DOCLINE version 1.0 went live to all National Network of Libraries of Medicine (NN/LM) libraries. DOCLINE is NLM's online interlibrary loan request routing and referral system. DOCLINE provides four main functions to its 3,000 participating U.S. and Canadian medical libraries:

- DOCUSER—provides directory and interlibrary loan information on NN/LM participating libraries.
- REQUESTS—allows users to make document requests which are routed automatically to libraries who report owning the specific year or volume requested.
- SERHOLD—provides journal holdings information.
- Loansome Doc Patron Administration—allows libraries to maintain administrative information on their Loansome Doc users.

The new web-based DOCLINE system is a key step in NLM's Systems Reinvention. Developed at NLM, the system interfaces seamlessly with other NLM products and services including PubMed and the online catalog LOCATOR*plus*.

The new DOCLINE replaced two legacy systems, DOCLINE and SERLINE. Closing down the mainframe DOCLINE system was the end of an era. Since 1985, DOCLINE has been running on a mainframe computer at NLM. On July 14, 2000, the Borrow function of the

current mainframe based DOCLINE system was permanently removed. It had handled more than 10,000 ILL transactions per day at no cost. Similarly, SERLINE served many functions over the last 30 years. It was an authority file for serials data and was also the link between SERHOLD and MEDLINE.

Data Creation And Maintenance System

The new web-based Data Creation and Maintenance System (DCMS) indexed its first record successfully this year. This was a big step forward in replacing the legacy Automated Indexing Management System (AIMS). Eventually, DCMS will support all journal articles that are indexed with MeSH headings. By the end of September 2000, all data (keyboarded, scanned/OCR, electronic) are being received in XML.

The DCMS will provide all current mainframe-based functionality as well as various enhancements. It provides citation creation, online journal assignment, journal tracking function, SGML article review, SGML issue verification, and citation maintenance functions. The DCMS has interfaces with several other systems, including Mesh2000, ILS, and PubMed. NLM maintains a comprehensive database of medical citations, including MEDLINE records and non-MEDLINE records. In combination, these modern NLM reinvented systems support:

- Making data available in FTP & XML;
- Leveraging the availability of journal articles in electronic form;
- Streamlining the processes used produce and/or provide information to customers; and
- Redistributing data to NLM's reinvented systems.

Currently, the phase I version of the DCMS application has been completed. Limited production use of DCMS for data creation started in July 2000. The process to extract XML data from the DCMS database was designed and implemented. This will be used to send the released records to the mainframe to be merged with records coming from AIMS. The AIMS

database is housed on an IBM mainframe, which is targeted for removal in 2001.

Detailed requirements and design of data creation workflow portion of Phase II of the application has been completed. The requirements and design of the citation maintenance portion of Phase II of the DCMS application also was begun. This portion will include maintenance of individual citations, batch updates and MeSH changes. In addition, analyzing the user requirement for DCMS reports has been started. OCCS staff worked with BSD staff to identify the report format, priority list, and DCMS Impromptu catalog.

While DCMS is being developed, OCCS staff still maintain AIMS, the legacy system that is used by in-house and contract indexers, revisers and others to prepare citations from medical journals for entry into Index Medicus and MEDLINE. OCCS personnel provide technical guidance and function as the primary point of contact for the system.

Use of XML and DTD for Creating MEDLINE Data

As part of the Systems Reinvention, XML (eXtensible Markup Language) was chosen this year to be the new tagged format for disseminating MEDLINE bibliographic citation data. A DTD (Document Type Definition) developed by OCCS in close coordination with other parts of NLM is being used to define the structure of this XML. This XML standard will be the only distribution format for MEDLINE data created for the 2001 indexing year.

This decision strengthens NLM's commitment to distribute its journal citation data in a format that is widely described and, therefore, familiar to many in the information industry, especially in the Internet Web environment. Choosing XML as the data format was a natural extension of NLM's use of XML to receive bibliographic data electronically from publishers.

Selection of the XML standard is one of the key technical underpinnings of System Reinvention. The new web-based Data Creation and Maintenance System indexed its first record successfully this year. This was a big step forward in replacing the legacy system AIMS,

and moving the MEDLINE database from a proprietary database on an IBM mainframe to a commercially available relational database product on Unix.

Serials Extract Database

OCCS developers designed the Serials Extract Database to support a number of NLM functions that the Voyager ILS does not support efficiently because of its proprietary nature. The DCMS development team designed programs to extract serial information from the ILS and create a database specifically designed to support NLM functions to include the DCMS, LSTRC, DOCLINE, and Publications, and to export journal authority data to PubMed.

Synchronization of the two databases is controlled by scripts that periodically update the Serials Extract Database. Manual updating to the Serials Extract Database is not supported. The automatic updating of the Serials Extract database was placed in production during the last quarter. OCCS staff continue to develop and modify the Serials Extract Database to meet the needs of applications with which it will interface. Such enhancements include modifications to the ISSN as well as the Citation Subset and Indexer Subset changes. The Citation Subset is used to support publishing various indexes such as Index Medicus and the Nursing Index. Each citation may be in one or more Citation Subset. The Indexer Subset is used to produce counts of articles indexed by the Indexer Subset.

Streamlined Data Distribution

Another result of the Systems Reinvention effort is the streamlining of data distribution. Besides the use of the Serials Extract database noted above, the year-end processing of MEDLINE data is also being improved. The goal for the future is to rely on Internet and advanced tape technology instead of on magnetic tapes for distribution of journal citation data. Already, the first licensee has cancelled the request for magnetic tapes for MEDLINE, preferring to use FTP. The goal is to have the updates be transmitted in XML format via FTP over the Internet and to use DLT tape

technology for the full file. Licensees have been advised that the final MEDLINE tapes in the legacy ELHILL format will be distributed after the final weekly update for the 2000 production year, which is scheduled for October 21, 2000.

OCCS staff are still maintaining legacy systems that are not yet off the mainframe. Thus, performing year-end processing and maintaining comprehensive databases requires some combination of data from reinvented systems and legacy systems. This is a complicated effort. While OCCS is transitioning the old MEDLARS systems off the mainframe, it is important to note that the goal is not so much to get off the mainframe as it is to have more modern systems that are easier to maintain and use, easier to link to, and more flexible.

List of Journals Indexed and List of Serials Indexed

For the first time this year, OCCS completed the *List of Journals Indexed* (LJI) and *List of Serials Indexed* (LSI) publications using the Serials Extract Database. They were completed, approved by LO, and sent to the publishers. The publication is now in hard copy and has been distributed. For the first time this year, these indexes were available for web download. They are available online in Portable Document Format (PDF), which require the use of the Adobe Acrobat Reader, which can be downloaded from Adobe's Web site at no charge. Files are available also in DOS text format.

MEDLINEplus Consumer Health Information

This year, OCCS and LO introduced a number new features in MEDLINEplus, including support for the medlineplus.gov URL, the addition of clinical trials data, the addition of a medical encyclopedia, the addition of extensive drug information based on the United States Pharmacopeia, and the introduction of a new design and organizational structure. There was also the addition of a FAQs and a welcoming video from the NLM director.

The neatly crafted MEDLINEplus web pages require much behind-the-scene work. For example, the aggressive schedule for version 4.7

(for the drug information) dictated the use of HTML. This required a number of perl scripts to parse the input data into separate subdirectories for each drug monograph, several SQL scripts to populate tables to create the A-Z breakdown, and cross-references between brand names and monograph names. In addition, several ColdFusion modules were needed to create the drug information page.

In addition to the new releases, the MEDLINEplus team continued to support production requests and to work on the prototype of the link checker to add functionality and robustness. New methods were created for the development and testing of MEDLINEplus under a new ColdFusion development environment. The development team also evaluated existing MEDLINEplus code with a vision toward a future infrastructure and a GUI redesign. A process for updating the clinical trials information and for providing that information to the clinical trials web team was also developed this year.

Search Engine

OCCS is continually refining the search engine supporting the NLM web site to improve its search capabilities. NLM uses ht://Dig as a search engine for the main NLM site. This was selected by OCCS, with advice and assistance from LO, after running it against a number of benchmark tests. OCCS is constantly monitoring and refining the search engine, as well as periodically reviewing the capabilities of engines other than ht://Dig. Currently, the team is working with CIT on a spell checker from Wintertree for use in improving search capabilities. The current plan is to implement Wintertree first for MEDLINEplus, then to apply it to the NLM main web site.

Reporting System (Active Concepts' FunnelWeb)

The OCCS web team supports and maintains all daily, monthly, and quarterly log reports for the NLM main web site and MEDLINEplus. But better analysis and quality reports were needed to meet the requirements of LO, web contributors, and web content group. The OCCS web team evaluated 16 different web

log analyzer packages, rating them in terms of report capability, output quality, interface flexibility, and speed. Active Concepts' FunnelWeb Pro was selected. Key features include streaming media analyses, cluster analysis, proxy analysis, support for virtual domains, click-stream analysis, incremental log analysis, remote administration via web, and online advertising analysis.

Web Content Management Software

The OCCS Webmaster team has investigated several web content management software packages. In order to ensure the integrity of the NLM web site, creation and deployment of web pages and related application files should be performed in a controlled environment. The Teamsite package is one possible solution that has been proposed by OCCS and a demonstration was arranged. While the demonstration went well, a number of additional requirements were identified. The OCCS team is experimenting with other software and will prepare a report comparing the various packages.

NLM Technical Bulletin

OCCS received a request from the Medlars Management Section of LO to allow users to print the entire issue of the *NLM Technical Bulletin* instead of only one article at a time. OCCS worked with MMS representatives to create a new template to support a "print all" function. OCCS developed and tested a program to perform this function and MMS will be starting user acceptance testing next year.

Administrative Support Systems

This year, OCCS continued to increase its development support for its internal customers at NLM, working on four administrative systems: inventory control, personnel policy, administrative manual, and online request for service.

OAM Inventory Control Project

The OAM Inventory Control System allows NLM users to order office supplies online and assists OAM in inventory management. This year the OCCS Inventory Control project team made a number of technical enhancements such as adding the capability to recall a filled order and make changes, including accepting returned items. The functionality of the system also was enhanced. Customers now are notified by e-mail about back orders or when new stock items arrive. Several additional screens now include the unit prices of items and identify the shipping/tracking numbers for stock received. A major enhancement this year was the addition of pictures in an online image catalog function. So far, approximately 600 pictures of inventory items have been taken. OCCS is reviewing the quality of pictures and their corresponding labels.

Personnel Administrative Control Project

OCCS is developing an online Personnel Administrative system that will track personnel information. The system will include employee information, recruitment actions, personnel actions, and award information. The project was developed in ColdFusion with Impromptu reports. All of the major functions have been implemented and tested. User acceptance testing by Personnel is now in progress.

NLM Manual Chapters And Delegations Of Authority

The new online version of NLM Administrative Manual was made available to staff on the NLM Intranet in September 2000. The manual is divided into four sections: Manual Chapters, Delegations of Authority, Functional Statements, and Organizational Charts. The OCCS development team and NLM Personnel Office staff converted more than 80 documents from various formats, including MS Word and HTML, to the PDF format of Adobe Acrobat. Currently, the OCCS development

team is testing the search function and working on the future maintenance procedures with Personnel.

OAMS Request for Service System

Following the successful implementation of the OAM Inventory Control System, NLM's Office of Administrative Management Services (OAMS) requested that

OCCS develop an automated online system for receiving and tracking requests for service. Currently, requests for service are tracked via a paper trail. The services covered by the proposed system include maintenance trouble calls, telecommunications work/trouble calls, and transportation and messenger service. The same OCCS project team that implemented Inventory Control system will work on this project.

ADMINISTRATION

Donald C. Poppke
Associate Director for Administrative
Management

National Performance Review

The NLM System Reinvention is a high-priority initiative conducted by NLM in support of its role as a reinvention laboratory under the National Performance Review. The project is designed to reinvent the Library's information systems, to move to a more flexible, powerful, and maintainable computer system that will improve internal processing and provide innovative services to outside users. Significant progress in system reinvention was made in several areas in FY2000:

Integrated Library System: NLM acquired and installed Voyager, an integrated library system (ILS) in FY1999. The focus of FY2000 was to develop seamless interfaces between Voyager and other NLM systems. These include NLM's system for processing interlibrary loan requests (Relais), NLM's document delivery system (DOCLINE), NLM's retrieval system (PubMed), NLM's online public access catalog (LOCATORplus), NLM's system for data creation and maintenance of MEDLINE articles (DCMS), and the NLM Gateway.

PubMed Retrieval System: PubMed is a World Wide Web retrieval service developed by NLM that provides access, free of charge, to MEDLINE. As part of the System Reinvention initiative, the MEDLINE database in PubMed was expanded to include the journal citations that have been in the HealthStar database. PubMed also contains links to the full-text versions of articles at participating publishers' Web sites. In addition, PubMed provides access and links to the integrated molecular biology databases maintained by the NCBI. These databases contain DNA and protein sequences, genome mapping data, and 3-D protein structures. PubMed has been widely accepted by the biomedical community and consumers.

Document Delivery: DOCLINE is the Library's automated interlibrary loan (ILL) request and routing and referral system. The purpose of this system is to provide improved document delivery service among libraries in the NN/LM. This improved system efficiently links and routes electronic journal holdings data from potential lending libraries to the borrower. NLM developed a new web-based application, replacing DOCLINE, which has served the biomedical community for over 15 years. The new system provides one-stop shopping for the three major components of DOCLINE, including: 1) DOCUSER7—user institutional information, including address, contact names, interlibrary loan services, NN/LM membership information, and routing tables; 2) SERHOLD7—journal holdings information; and 3) REQUESTS, the functions of Borrow, Lend and Status/Cancel. DOCLINE seamlessly interfaces with the Voyager ILS and NLM's journal citation retrieval engine. DOCLINE currently supports 3 million interlibrary loan requests annually.

Data Creation and Maintenance: The new indexing Data Creation and Maintenance System (DCMS) replaced several legacy systems used for online indexing and editing of bibliographic citations for MEDLINE and related files. Key design features are the use of the World Wide Web, a relational database, and the Extensible Markup Language (XML). More than 425,000 citations from approximately 4,400 journals are added to MEDLINE annually. Citation data are created by one of three separate mechanisms: (1) from SGML-tagged data that is submitted to NLM by journal publishers; (2) from a scanning/optical character recognition (OCR) operation; or (3) from a manual keyboarding operation.

User Access Services: Three major efforts were undertaken to ensure a smooth transition while System Reinvention was under way. These efforts were the streamlining of NLM's Customer service activities, the modernization of NLM's distribution of databases (including MEDLINE) to external organizations, and the gateway services.

Customer service—Committed to providing excellent customer service for the growing number of products and services provided by the Library, NLM established a centralized Customer Service function in the Reference and Customer Services Section. The goals of this reinvention project were to:

- improve the quality and timeliness of service to the customer;
- increase NLM staff productivity; and
- turn customer feedback and staff knowledge into product improvement.

A centralized telephone service number, 1-888-FINDNLM, gives quick access to a customer service staff member. Customer Q7 customer service software from Quintus Corporation was installed to handle all customer inquiries received by telephone, electronic mail, fax and postal mail. Over 60,000 questions are handled annually. The system enables staff to efficiently assign questions to customer service staff, track timeliness, conduct quality reviews, and quickly identify product problems so that they can be fixed.

Modernization of NLM's Distribution of Databases—NLM exports completed MEDLINE and other database citations to a number of external organizations. Under the legacy NLM systems licensees were required to meet an NLM defined standard for data representation, and distribution by tape media of the full MEDLINE file required shipping more than 140 tapes. This entire process has been replaced with industry standard data representation utilizing the eXtensible Markup Language (XML) and UNICODE. The full MEDLINE file is now available for distribution on a single tape. Subsets of MEDLINE are available from the NLM FTP site.

Gateway services—The NLM Gateway presents a single interface that lets users search simultaneously in multiple NLM retrieval systems. Its target audience is the Internet user who comes to NLM not knowing exactly what is here or how best to search for it. Key features include providing “first-stop shopping” for an increasing number of NLM's information resources, including citations, full text, video,

audio, and images, and use of NLM's Unified Medical Language System (UMLS) to help users find effective search terms. The NLM Gateway presently provides access to a number of NLM resources including PubMed, LOCATORplus, and MEDLINEplus. Future releases will add additional resources as development continues. Over time the NLM Gateway will replace the Internet Grateful Med system.

Financial Resources

In FY 2000, the Library had a total appropriation of \$214,068,000. Table 11 displays the FY 2000 authority plus reimbursements from other agencies.

Table 11

Financial Resources and Allocations, FY 2000

Budget Allocation:

Extramural Programs	\$44,833
Intramural Programs	160,103
Library Operations	(67,838)
Lister Hill National Center for Biomedical Communications.....	(48,692)
National Center for Biotechnology Information	(33,713)
Toxicology Information	(9,860)
Research Management and Support.....	9,132
Total Appropriation	214,068
Plus: Reimbursements	14,743

Total Resources228,811

The 2000 appropriation language authorized the Library to use personal services contracts and provided for the availability of \$4.0 million without fiscal year limitations. These authorities are key elements of NLM's system reinvention initiative.

Personnel

In October 1999, **Mr. Anthony Y. Tse** joined the staff of LHCBC as a Research Fellow. Mr. Tse is completing his Ph.D. with the University of Maryland, College of Library and

Information Services. At Lister Hill Center, Mr. Tse will work with the Clinical Trials Data Base and the Natural Language Processing research groups.

In October 1999, **Ms. Elizabeth A. Pope** joined the staff of NCBI as a Staff Scientist. Ms. Pope received her Bachelors degree in biology in 1979 from the University of California, San Diego. Ms. Pope will be an integral player in the PubMed Central project. She will work with publishers, societies and organizations submitting the data to insure that the PubMed Central is a success.

In November 1999, **Suzanne T. Szak, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Post-Doctoral IRTA. Dr. Szak received her Ph.D. in biochemistry from Vanderbilt University School of Medicine, Nashville, Tennessee in 1999. Dr. Szak will work on databases and software tools to aid in the interpretation of gene expression profiling data.

In December 1999, **Ms. Becky Lyon** was selected for the position of Deputy Associate Director, Division of Library Operations. A former NLM Associate, Ms. Lyon held positions in the Technical Services Division, where she was instrumental in the original automation of book acquisitions functions, and in the Lister Hill Center, where she managed an experimental computer-assisted training network. Ms. Lyon has led the National Network of Libraries of Medicine (NN/LM) program, playing a major role in expanding NLM's outreach initiatives, including recent efforts to reach the general public. Ms. Lyon also assisted with the national deployment of DOCLINE, directed efforts to increase Internet connectivity among hospital libraries, and directed the initial study that led to the recent consolidation of first line customer service in LO.

In December 1999, **Mr. Aravind L. Iyer**, joined the staff of the Computational Biology Branch of NCBI as a Staff Scientist (VP). Mr. Iyer received his Masters in biotechnology in 1995 from the University of Pune (India), after which he joined the Ph.D. program at Texas A&M University, College Station, Texas. He is currently completing the final requirement for his Ph.D. in genetics. Mr.

Iyer will be working on the development of a new generation of protein databases that combine structural and evolutionary information for the purpose of robust classification of proteins and reliable prediction of their activities and functions.

In December 1999, **Svetlana A. Shabalina, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Research Fellow. Dr. Shabalina received her Ph.D. degree in molecular biology from the Institute of Molecular Biology, Russian Academy of Sciences, Moscow in 1994. Dr. Shabalina will take advantage of the rapidly growing amount of data on related eukaryotic genomes in order to study the extent and nature of functional constraint in various regulatory regions.

In December 1999, **Ian J. Harrison, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Research Fellow. Dr. Harrison received his Ph.D. in biology from the University of Bristol, U.K., in 1987. His research has focused primarily on the taxonomy and classification of fishes. Dr. Harrison's research has resulted in 14 scientific publications in international journals and 8 publications in popular journals. Dr. Harrison will be involved in the NCBI/GenBank taxonomy project.

In January 2000, **Jane M. Carlton, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Research Fellow (VP). Dr. Carlton received her Ph.D., in molecular genetics in 1995 from the University of Edinburgh, Scotland. Dr. Carlton will complement and strengthen the existing NCBI research efforts on malaria which employ computer analyses of comprehensive genetic and genomic data, in close collaboration with the NIAID intramural laboratories and genome sequencing centers.

In January 2000, **Mr. Grigoriy Starchenko** was appointed as Staff Scientist with the Information Engineering Branch, NCBI. Mr. Starchenko received a Master's degree in mechanical engineering from Moscow State Technical University in 1990. For almost five years, while working as a contractor, Mr. Starchenko has been actively involved in creating and maintaining the databases and

software for the PubMed project and the Entrez retrieval system. In his new position, Mr. Starchenko will continue his outstanding stewardship of PubMed.

In February 2000, **Mr. James Marcetich** was selected as the new Head of NLM's Index Section, Bibliographic Services Division. As a unit head in the Index Section for the past 12 years, Mr. Marcetich has supervised the work of indexers and revisers. His experience also includes coordinating the online indexing system used by approximately 100 in-house, contract, and foreign center indexers; serving as the Section's lead on the development of the Library's new indexing and maintenance system; and serving as alternate project officer for NLM's indexing contracts. Mr. Marcetich came to NLM in 1979 as a library associate in NLM's postgraduate training program and before becoming a unit head, was an indexer and reviser for eight years.

In February 2000, **Mr. Garry Fox** was appointed Chief, Systems Support Section, Systems Technology Branch, Office of Computer and Communications Systems (OCCS). Mr. Fox received his BS in computer science from West Virginia Institute of Technology, Montgomery, West Virginia. In 1991, Mr. Fox joined the NLM as a Computer Systems Programmer with the Systems Support Branch, OCCS. In his new position, Mr. Fox will serve as the technical expert for overseeing all activities ensuring that NLM computer equipment, operating systems, languages, and associated software are managed and utilized effectively.

In March 2000, **Mr. David Kenton**, joined the staff of the Information Engineering Branch, National Center for Biotechnology Information as a Staff Scientist. Mr. Kenton received his BA in mathematics from the University of Connecticut in 1964. In 1972, he joined the OCCS staff as one of the original 3-member team that designed, built, and deployed the ELHILL information retrieval system. ELHILL has been the central resource for building and retrieving MEDLINE records for more than 25 years and Mr. Kenton has been a key player in maintaining and expanding the capabilities of the system. Mr. Kenton has also voluntarily taken a very active role in making

sure that the new PubMed system, developed at NCBI, interfaced smoothly with the ELHILL system. With his move to NCBI, Mr. Kenton will contribute substantially to the continued growth and development of these systems.

In March 2000, **Ms. Kathy Cooper** joined NLM as Chief, Desktop Services Section, Systems Technology Branch, OCCS. Ms. Cooper received her BS in technology and management from the University of Maryland, College Park, Maryland. At OCCS, Ms. Cooper will be responsible for ensuring that all NLM divisions receive effective and timely advice, guidance, and technical support for their personal computers and workstations. She will also be responsible for planning, organizing, and directing studies to identify and evaluate alternatives and to determine the feasibility of implementing new procedures and techniques to meet the software and desktop needs of the organization.

In March 2000, **Gabor T. Marth, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Staff Scientist. Dr. Marth received his doctorate in systems science and mathematics from Washington University, St. Louis in 1994. At NCBI, it is anticipated that Dr. Marth will continue his work on PolyBayes, focusing on the theoretical and algorithmic improvement of sequence-based SNP (single nucleotide polymorphisms) detection. He is expected to help NCBI develop a standard for reporting the quality of computationally predicted SNPs.

In March 2000, **Karen L. Clark, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Clark earned her Ph.D. in biological chemistry from the University of Michigan in 1985. In 1997, Dr. Clark began working for ComputerCraft Corporation as a GenBank indexer. Since that time, she has been involved in the processing, annotation, and updating of direct submissions to GenBank. Dr. Clark is expected to use her experience to continue to excel in her tasks in order to maintain the integrity of the GenBank database while striving to increase general indexing efficiency and quality.

In March 2000, **Linda K. Yankie, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist.

Dr. Yankie received her Ph.D. in biochemistry from the University of Maryland in 1990. In 1997, Dr. Yankie joined ComputerCraft Corporation as a Scientific Data Analyst for GenBank working on-site at NCBI. Dr. Yankie is expected to use her experience to continue to excel in her tasks in order to maintain the integrity of the GenBank database while striving to increase general indexing efficiency and quality.

In March 2000, **Susan L. Schafer, Ph.D.**, was appointed to the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. Schafer received her Ph.D. in biochemistry from the University of Maryland in 1993. In 1998, Dr. Schafer joined ComputerCraft Corporation as a Scientific Data Analyst for GenBank. Dr. Schafer will continue to participate in all phases of processing direct submissions to GenBank and will also assist in other special projects such as the testing of software, both for use by indexers and users of GenBank.

In March 2000, **Jennifer C. McDowell, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI as a Staff Scientist. Dr. McDowell received her Ph.D. in biochemistry from Cornell University in 1994. In 1997, Dr. McDowell began working for ComputerCraft Corporation as a GenBank indexer. Since that time, she has been involved in initial processing, the annotation, and updating of direct submissions to GenBank. Dr. McDowell will continue to participate in all phases of processing direct submissions to GenBank.

In March 2000, **Irving K. Jordan, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as an IRTA Fellow. Dr. Jordan received his Ph.D. in genetics from the University of Georgia in 1998. At the NCBI, Dr. Jordan will perform research on phylogenetic classification of proteins and evolutionary and functional analysis of protein super families using methods of computational biology. This research direction is critical for meaningfully interpreting the rapidly accumulating genomic data and incorporating the acquired information into molecular biology databases. Dr. Jordan is expected to continue making major contributions to evolutionary genomics.

In March 2000, **Yonil Park, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Fogarty Visiting Fellow. Dr. Park received her Ph.D. in applied mathematics from the Korea Advanced Institute of Science and Technology in February 2000. She will conduct research by developing analytic upper bounds for the statistical significance of gapped alignments. The development of fast methods for computing statistical significance for gapped alignments would aid the effectiveness of the BLAST programs, a major software service at the NCBI.

In March 2000, **Igor B. Rogozin, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Research Fellow (VP). Dr. Rogozin received his Ph.D. in biology from the Institute of Organic Chemistry of the USSR Academy of Sciences (Siberian Branch) in 1990. At the NCBI, Dr. Rogozin will perform research on comparative genome analysis and evolution of operons, regulons and large-scale genome organization in prokaryotes using computational biological methods. This research direction is critical for meaningfully interpreting the rapidly accumulating genomic data and incorporating the acquired information into molecular biology databases.

In March 2000, **Vivek Anantharaman, Ph.D.**, joined the staff of the Computational Biology Branch of NCBI as a Visiting Fellow in the NIH Visiting Program. Dr. Anantharaman, a native of India, received his Ph.D. in biomedical science from Old Dominion University and Eastern Virginia Medical School in 1999. At the NCBI, Dr. Anantharaman will perform research on phylogenetic classification of protein and evolutionary and functional analysis of protein superfamilies using methods of Computational Biology.

In April 2000, **Eugene P. Yaschenko**, joined the staff of the Information Engineering Branch of NCBI as a Staff Scientist. Mr. Yaschenko received a High Education Diploma in Physics and Materials Science from Moscow State University in 1992. He received his master's degree in physics from the Catholic University in Washington, D.C. Mr. Yaschenko came to NCBI in 1995 as a contract software developer. He worked on a team doing a major redesign of the NCBI sequence databases. Mr.

Yaschenko will continue to serve as the focal point for testing new database approaches. He is the acknowledged expert on technical relational database and client/server issues.

In April 2000, **John J. Anderson, Ph.D.**, joined the staff of the Computational Biology Branch of NCBI on an Intramural Research Training Award. Dr. Anderson received his Ph.D. this year in molecular and cellular biology from the University of Arizona in Tucson. At NCBI he will work on computational problems related to the understanding of gene regulatory events in eukaryotes. He will also develop new methods for the identification of genetic networks in this genome.

In May 2000, **Benjamin A. Shoemaker, Ph.D.**, joined the staff of the Computational Biology Branch of NCBI on an Intramural Research Training Award. Dr. Shoemaker received his Ph.D. in computational chemistry from the University of Illinois at Urbana-Champaign, September 1999. At NCBI, Dr. Shoemaker will join the CDD project, a "Conserved Domain Database." The work will involve research in sequence/structure alignment algorithms, to support automated updates of multiple alignments describing protein domains.

In May 2000, **Liora Z. Strichman-Almashanu, Ph.D.**, joined the staff of the Computational Biology Branch of NCBI as a Fogarty Visiting Fellow. Dr. Strichman-Almashanu, a native of Israel, received her Ph.D. this year in human genetics from the Johns Hopkins University School of Medicine. At NCBI, Dr. Strichman-Almashanu will concentrate on a project dealing with processed pseudogenes. Processed pseudogenes are intronless, mRNA-like sequences, which have been derived from actively transcribed genes by a process of reverse transcription and integration into new genomic locations.

In May 2000, **Simon Y. Liu, Ph.D.**, was appointed to the Senior Executive Service position of Director, Information Systems, NLM. Dr. Liu serves in a dual staff and line management role in the broad field of information systems. In a staff capacity, Dr. Liu advises the Director and Deputy Director about computer and communications technology affecting library and information activities. In

his line management role, as Director of the Office of Computer and Communications Systems (OCCS), Dr. Liu directs the activities of OCCS. Dr. Liu has over 14 years of professional experience in planning, developing, and managing Information Technology (IT) programs in both the Federal Government and private industry. His educational background includes an MBA in business administration from University of Maryland (May 1999), and Ph.D. in computer science from George Washington University (August 1995).

In May 2000, **Jean Thierry-Mieg, D.Sc.**, joined the staff of the Information Engineering Branch, NCBI as a Senior Research Fellow (VP). Dr. Thierry-Mieg received his D.Sc. degree in Theoretical Physics from Universite de Paris, France in 1978. He spent several extended periods of time as a visiting scientist at some of the leading research centers on theoretical physics worldwide, including the California Institute of Technology (CalTech), Harvard University, and the Lawrence Berkeley Laboratory in California. For his work on string theory in theoretical physics, he received a Silver Medal from the Centre National de la Recherche Scientifique in France. Dr. Thierry-Mieg's experience analyzing the nematode genome sequence will be invaluable to the mandate of the NCBI in giving comprehensive database access to a fully annotated version of the data. The NCBI has also been charged with the assembly of the large contig data into coherent sequences. Dr. Thierry-Mieg has also had significant experience in these problems with the smaller *C. elegans* chromosomes. He will work on a project to make the same kind of dramatic improvements to the human ESTs and genomic sequence.

In May 2000, **Danielle B. Thierry-Mieg, D.Sc.**, joined the staff of the Information Engineering Branch, NCBI as a Senior Research Fellow (VP). Dr. Thierry-Mieg received her D.Sc. degree in genetics from Universite de Paris, France in 1984, where she worked on the developmental genetics in *Drosophila melanogaster*. Since 1993, Dr. Thierry-Mieg has been involved in many bioinformatics research projects particularly as it is applied to the genome of *Caenorhabditis elegans*. She has spent several research and training periods at

some of the most respected bioinformatics institutions worldwide. Some of Dr. Thierry-Mieg's most significant contributions have been in the area of computational biology on the genome of *C. elegans*. She will complement and strengthen the existing NCBI research efforts on the assembly and annotation of the human genome sequence, which employs computer analyses of comprehensive genetic and genomic data.

In June 2000, **Angela B. Ruffin, Ph.D.**, was appointed Head, National Network of Libraries of Medicine (NN/LM) Office in the Division of Library Operations. Dr. Ruffin has 10 years of successful experience in coordinating outreach programs for the NN/LM Office, starting with the first round of Grateful Med outreach projects. Prior to coming to NLM in 1990, Dr. Ruffin taught at several Schools of Library and Information Science and served as media coordinator for the Durham City Schools. Dr. Ruffin received her B.A. from Spelman College, her M.S.L.S. from Atlanta University (now Clark-Atlanta University), and her Ed.M in educational psychology from Boston University. She received her Ph.D. in Information and Library Science from the University of North Carolina, Chapel Hill.

In June 2000, **Mr. Dwight H. Mowery, Jr.**, was appointed NLM's new Grants Management Officer. Mr. Mowery joined the NLM in 1995 and he has been performing grants management activities for the NIH community since 1987. Mr. Mowery is responsible for the overall fiscal and administrative management of the Library's grants programs including research grants, resource grants, publications grants, training grants, interagency and inter-institutional agreements, and complex cooperative agreement mechanisms. Mr. Mowery received a bachelor's degree in sociology from St. Francis College, Pennsylvania.

In June 2000, **Simon Baatz, Ph.D.**, joined the staff of the History of Medicine, Division of Library Operations as a Special Expert. Dr. Baatz, a native of Great Britain, received his Ph.D. in history and sociology of science from the University of Pennsylvania in 1986. From 1981 to 1986, Dr. Baatz has held teaching positions at the University of

Pennsylvania, University of Sussex, and University of Exeter. While at the History of Medicine Division, Dr. Baatz will research and write a new history of the National Library of Medicine, in one volume. The focus of the new history is to be the development and growth of the Library's programs in the postwar period, especially in the last 30 years.

In July 2000, **Ms. Martha R. Szczur** was appointed as Special Expert in the Division of Specialized Information Services (SIS). Ms. Szczur will serve as Deputy Director, responsible for information technology and systems development and for the reinvention of information systems developed in direct support of SIS programs. Her educational background includes a B.A. from Converse College, Spartanburg, SC (1968) in mathematics. Ms. Szczur has over 30 years of diverse experience managing and implementing complex, large-scale information management systems and technology programs, spending the last 20 at NASA/Goddard Space Flight Center (GSFC). At NASA she served as Chief of the Information Systems Center, Applied Engineering and Technology Directorate.

In July 2000, **Donald C. Comeau, Ph.D.**, joined the staff of the Computational Biology Branch of NCBI as a Research Fellow. Dr. Comeau received his Ph.D. in theoretical chemistry in 1990 from Ohio State University. In 1993, Dr. Comeau accepted a position as Associate Professor of Computer Science at Columbia Union College. As a Research fellow, Dr. Comeau will work on various aspects of the electronic textbook project that is currently in its early developmental stages at NCBI. He will be responsible for developing improved software and methods to process the textbooks for links to PubMed documents.

In July 2000, **Mary Moore, Ph.D.**, joined NLM under an Intergovernmental Personnel Assignment (IPA) agreement. Dr. Moore is currently the Dean of Library and Information Resources and Professor, College of Communications at Arkansas State University (ASU). At NLM, Dr. Moore will serve as coordinator of the Associate Fellowship program, NLM's post-masters training and internship program for health sciences librarians. She will also advise NLM on matters related to

distance education and the role of health sciences librarians in telemedicine programs. Dr. Moore has a B.A. and M.A. in library and information sciences from the University of Missouri, Columbia. She received her Ph.D. in library and information science from the University of Texas at Austin.

In July 2000, **Vicky Siu-Ngan Choi**, joined the staff of the Informational Engineering Branch of NCBI in July as a pre-doctoral Visiting Fellow. Ms. Choi, a native of Hong Kong, is presently a graduate student in the computer science department at Rutgers University. Her interests are computational molecular biology, design and analysis of algorithms. At NCBI, Ms. Choi will use her training as a theoretical computer scientist and her understanding of real data from having worked in a genome lab to improve current heuristics for the contig layout problems.

In August 2000, **Lakshminarayan M.S. Iyer, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Visiting Fellow. Dr. Iyer, a native of India, received his Ph.D. in biology from the Department of Biology of Texas A&M University. At the NCBI, Dr. Iyer will perform research on comparative genome analysis and phylogenetic relationships between genomes using methods of computational biology. This research direction is critical for meaningfully interpreting the rapidly accumulating genomic data and incorporating the acquired information into molecular biological databases.

In August 2000, **Natsuhiko Futamura**, joined the staff of the Computational Biology Branch, NCBI as a pre-doctoral Visiting Fellow. Mr. Futamura, a native of Japan, is currently a doctoral student at the Department of Electrical and Computer Engineering at Iowa State University, Ames Iowa. Mr. Futamura received his M.S. degree in computer and information science in 1996 from Syracuse University. Mr. At NCBI, he will implement new modules for BLAST, which is the most important software package distributed by NCBI.

In September 2000, **David L. Wheeler, Ph.D.**, joined the staff of the Information Resources Branch, NCBI as a Staff Scientist. Dr. Wheeler received his Ph.D. in biochemistry in 1990 from Old Dominion University and

Eastern Virginia Medical School. He was awarded two postdoctoral fellowships at NICHD and NIDDK during which he produced approximately 15 scholarly publications, including two reviews and a book chapter. Prior to joining the staff of NCBI, Dr. Wheeler was employed as a Senior Scientist with the Kevric Company, and served as a contractor with NCBI. Dr. Wheeler has applied his programming expertise to the creation of several Perl scripts for use by the Service Desk staff, an innovative graphical environment in TCL/TK for the integration of these scripts, called SDTool, and a Java applet called "NCBI Pathfinder" which is designed to assist users to navigate NCBI's web resources.

In September 2000, **Peter S. Cooper, Ph.D.**, joined the staff of the Information Resources Branch, NCBI as a Staff Scientist. Dr. Cooper received his Ph.D. in marine science/environmental sciences from the College of William and Mary, Williamsburg, Virginia in 1996. He held a postdoctoral fellowship at the National Institute of Environmental Health Sciences from 1996 to 1998. Prior to joining NCBI, he worked with the Kevric Company as a contractor with NCBI. As a member of the NCBI, Dr. Cooper will continue to improve the quality of user service support services and expand the scientific outreach program.

In September 2000, **Sean Turner, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Senior Research Fellow. Dr. Turner received his Ph.D. in biology from the University of Santa Cruz, California in 1985. At the NCBI, Dr. Turner will work in the NCBI/GenBank taxonomy project. The taxonomy project maintains a taxonomy database that includes names and classification for every organism that has been subject to molecular scrutiny. Dr. Turner's particular task will be to curate the Bacterial and Archaea section of the taxonomy database, maintain contacts with outside taxonomy consultants, and the creation and maintenance of genetic code and taxonomy web pages.

In September 2000, **Richard M. Desper, Ph.D.**, joined the staff of the Computational Biology Branch, NCBI as a Research Fellow. Dr. Desper received his Ph.D. in mathematics from Rutgers University, New

Brunswick, New Jersey in 1998. Dr. Desper is a mathematician and computer scientist who has adapted his skills to work on cancer genetics problems. His work at the NCBI will include work on cancer modeling problems. Dr. Desper will work on developing mathematical analysis tools to add value to the collected data. He will also continue his research in new mathematical methods for constructing phylogenetic trees.

In September 2000, **Lynn M. Schriml, Ph.D.**, joined the staff of the Information Engineering Branch, NCBI as a Research Fellow. Dr. Schriml received her Ph.D. in biology in 1997 from the University of Ottawa, Ontario, Canada. Over the last year, she has worked at the NCBI as a contract employee through Management Systems Designers, Inc. At the NCBI, Dr. Schriml will play a coordinating role for genomic information resources devoted to the rat and zebrafish, two experimental model organisms that are of particular importance to NIH researchers.

In September 2000, **Barton W. Trawick**, joined the staff of the Information Engineering Branch, NCBI as a pre-doctoral IRTA Fellow. Mr. Trawick is pursuing a Ph.D. in immunology from the University of Texas Health Science Center, where he received his M.S. in biochemistry. While at the University of Texas, Mr. Trawick conducted doctoral research in the area of organ transplantation immunology. Mr. Trawick will be working on the development of content-driven interactive websites to show NCBI users how our resources might be used to help make biological discoveries.

In September 2000, **Joana S.C. Carneiro da Silva, Ph.D.**, joined the staff of Computational Biology Branch, NCBI as a Visiting Fellow. Dr. Carneiro da Silva, a native of Portugal, received her Ph.D. in genetics from the University of Arizona. At NCBI, Dr. Carneiro da Silva will concentrate on mammalian non-coding regions. The project will be involved with screening human and murine genomes for domesticated copies of transposable elements, i.e. the copies that perform a function useful for the host.

Retirements and Resignations

In December 1999, **Ms. Sally Burke**, Deputy Executive Officer, NLM retired from the Federal government after 25 years of service. Ms. Burke began her career with the National Library of Medicine in 1974. During her tenure she served in various progressively responsible positions within the Office of Administration, Office of the Director, NLM. In 1990, Ms. Burke became NLM's Deputy Executive Officer. In this capacity, she shared responsibility with the Associate Director for Administrative Management in managing and directing the daily operations of budget, personnel, acquisitions, management analysis, and administrative services branches of the NLM.

In January 2000, **Ms. Nancy Roderer** departed the NLM to become Head of the Welch Medical Library at Johns Hopkins University. Prior to her departure, Ms. Roderer was a Technical Information Specialist with the Division of Library Operations and held the position as the NLM Library Associate Program Coordinator and Library Operations Research Advisor.

In February 2000, **Mr. John Seachrist, Jr.**, NLM's Grants Management Officer, Division of Extramural Programs, left the NLM to serve as Grants Management Officer for the National Center for Research Resources, NIH. While serving at NLM, Mr. Seachrist was responsible for the overall fiscal and administrative management of the Library's grants programs while ensuring overall compliance with applicable laws, regulations, policies, and procedures relating to grants management.

In February 2000, **Kenneth J. Address, Ph.D.**, resigned from his Staff Scientist position with the NCBI to join Xencor, Inc., Pasadena, CA as a senior computational scientist. Dr. Address began working with NCBI as an IRTA Fellow in 1997. He was converted to a Staff Scientist position with the Computational Biology Branch in 1998. Dr. Address's background is in experimental structure determination; during his employment with NCBI, he made critical contributions to the

design and implementation of NCBI's structure-structure similarity search service.

In March 2000, **Mark S. Boguski, M.D., Ph.D.**, departed the NLM to become Senior Vice President for Research at Rosetta Inpharmatics, Inc., Kirkland, Wash. Dr. Boguski joined the staff of NCBI as a Senior Staff Fellow in 1989. In 1998, he became a Senior Investigator in the Computational Biology Branch. Dr. Boguski's primary responsibilities included performing basic and applied research in computational biology and genomics. Dr. Boguski's accomplishments include cross-referencing of human genes with their counterparts in other organisms and contributing to the construction of a "transcript map" of the human genome using the sequences of ESTs; serving on the editorial advisory boards of *Science* and *Genome Research*; serving as a Member of the Genome Research Review Committee for the National Human Genome Research Institute; managing the EST Division of GenBank and co-organizing the Informatics session at the annual Cold Spring Harbor meeting on Genome Mapping and Sequencing.

In March 2000, **Ms. Frances E. Johnson** retired from the Federal government with over 33 years of service. She spent her entire Federal career at NLM, where she was initially appointed as a Librarian. In 1977, she became a Grants and Contracts Program Specialist with the Division of Extramural Programs. In this capacity, Ms. Johnson provided leadership and direction to grant and contract programs in support of major biomedical information programs. Serving in a lead role, Ms. Johnson monitored new developments, encouraged submission of proposals, and made recommendations in support of a balanced portfolio of programs and awards based on the priorities of the NLM and the needs of the field.

In July 2000, **Peter A. Clepper** retired from the Federal service with over 43 years of service in the Federal government. Mr. Clepper joined the NLM in 1968 as a Public Health Advisor in the Division of Extramural Programs. In 1979, he became a Grants and Contracts Program Specialist with responsibility for providing leadership and direction to NLM's research program project grants in the area of

informatics as applied to health care delivery and medical research. Information technology including computers and telecommunications, health science libraries, and training and education of informaticians were all included within the scope of the program.

In July 2000, **Ms. Patricia S. Page** resigned her position of senior Contract Specialist with the NLM, Office of Acquisitions Management to spend time with her family. Ms. Page began her career at NIH in 1972. She joined the staff of the NLM in 1984 where she served as a Team Leader and Contracting Officer for the management of R&D and Station Support contracts. She supported the acquisitions activities of various NLM Divisions, the NIH Center for Information Technology, and she conducted business reviews of NLM Interagency Agreements.

Special Milestone

In 1999, **James Cassedy, Ph.D.**, celebrated 50 years of service with the Federal government. Dr. Cassedy received his AB in American literature from Middlebury College, Vermont, and his Ph.D. in American civilization from Brown University, Rhode Island. He served in the United States Army from 1941 to 1946 and then spent 6 years with the United States Information Agency as Director of Cultural Centers in Haiti, Burma, and Pakistan. In 1968, Dr. Cassedy joined the NLM History of Medicine Division, where he has enjoyed an extraordinarily productive career as a scholar, bibliographer, mentor, and friend of historians of medicine throughout the country and around the world. At NLM, Dr. Cassedy has been Editor and Indexer of the *Bibliography of the History of Medicine* and of HISTLINE. He also organizes the History of Medicine Seminar series and continues a very active professional life. He has served as President of the American Association for the History of Medicine (1982-84) and received the NLM Regents Award in 1984. In addition to his numerous articles and reviews, Dr. Cassedy has published several books that have become classics in the field, including *Demography in Early America* (1969), *American Medicine and Statistical Thinking* (1984), and *Medicine in America* (1991).

Awards

The NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to Dr. Elizabeth Fee for her outstanding scholarship as an author and editor of highly regarded books, journals, and journal articles on the history of medicine and public health.

The Frank B. Rogers Award recognizes employees who have made significant contributions to the Library's fundamental operational programs and services. The recipient of the 2000 award was Mr. Kenneth Niles for dedicated leadership, ability to plan for the future, and creative use of technology in providing improved collection access services to the on-site patrons of the National Library of Medicine.

The NIH Director's Award was presented to one individual and two groups: Dr. Elliot Siegel for leadership of NLM's contributions to the Multilateral Initiative on Malaria and scientific communications in Africa; Ms. Joyce Backus, Mr. Joseph Hutchins, Ms. Lori Klein, Ms. Eve-Marie Lacroix, Ms. Wei Ma, Ms. Naomi Miller, and Ms. Robin Moore for their development of MEDLINEplus, which has substantially increased NIH's ability to deliver authoritative health information to the public; and Mr. Philip Nielsen, Mr. Rex Shuler, and Mr. Roy Standing in recognition of their outstanding guidance and technical competence ensuring that the NLM's critical computer systems were YEAR 2000 compliant.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to two employees: Ms. Cassandra Allen for important contributions as founding Chair of the NLM Diversity Council toward improving the NLM work environment and increasing opportunities for employees to develop; and Dr. Alexa McCray for her superb leadership in creating the Clinical Trials Database.

The NIH Merit Award was presented to four employees: Ms. Mary-Kate Dugan for exceptional achievement in managing the collections and serving as the Disaster

Prevention and Recovery Director for the National Library of Medicine; Ms. Marie Gallagher for technical leadership in the computer design and implementation of systems to make library collections available to the public through modern digital technology; Ms. Karen Riggs for exceptional management and effective administration of the small purchases function at the National Library of Medicine; and Ms. Julia Royall for leadership of NLM's contributions to the Multilateral Initiative on Malaria and scientific communications in Africa.

The NIH Quality of Work Life Award was presented to Dr. Stuart Nelson for genuine interest in the welfare of the staff as demonstrated by support of workplace solutions to help balance work/family lives; foster development of knowledge and workplace skills; and encourage and facilitate ergonomically healthy work habits and environment.

The Lifetime Achievement Award presented by the American Association for the History of Medicine was bestowed on Dr. James Cassidy as a long-time member of the American Association for the History of Medicine, in recognition of his distinguished record of support of the history of medicine.

The Phillip C. Coleman Award presented in recognition of significant contributions to the NLM by individuals who demonstrate outstanding ability to motivate colleagues. The recipient of the 2000 award was Ms. S. Margarita Ortiz for the initiative, enthusiasm, and invaluable assistance she provided while coordinating the Getting to Know NLM Presentations and for her role as a dedicated member of the NLM Diversity Council.

The NLM EEO Special Achievement Award was presented to Ms. Patricia Carson for the outstanding guidance and mentoring she provided to student employees, her invaluable assistance to the EEO program, and her involvement in the employee and EEO-related aspects of various highly successful NLM public activities and exhibits.

Table 12
FY 2000 Full-Time Equivalents

<i>Program</i>	<i>Full-Time Equivalents</i>
Office of the Director	11
Office of Health Information Programs Development	7
Office of Communications and Public Liaison.....	8
Office of Administration	56
Office of Computer and Communications Systems ..	64
Extramural Programs.....	15
Lister Hill National Center for Biomedical Communications.....	76
National Center for Biotechnology	79
Specialized Information Services.....	27
Library Operations	282
TOTAL FTEs.....	625

NLM Diversity Council

The NLM Diversity Council began the year by welcoming five new members: Nadine Benton, James Dean, Julian Owens, Tony Pirrone, and Julia Royall (reappointed). Each will serve a two year term from January 2000 through December 2001. Continuing on the Council are Cassandra Allen, Vivian Auld, Redmond Barnes, Dan Higgins, Sally Mooney, and Margarita Ortiz. The Council continues to receive support from its ex-officio members, Donald Poppke, David Nash, and Nadgy Roey. Julia Royall accepted the responsibilities of Council Chair and Vivian Auld became the Council Vice-Chair.

FY2000 Accomplishments:

- Continued to coordinate the NLM Director’s Employee Education Fund. In this second year of availability, the Fund enabled 25 staff to take 26 classes. Undergraduate classes made up 65 percent of the classes supported. The

- Diversity Council continues its efforts to publicize the availability of the Fund.
- Provided briefings on the activities of NLM’s Diversity Council or related issues to the Library of Congress, NIH Office of Equal Opportunity, and HHS Secretary’s Conference on Diversity.
 - In June the Council held the first session in the Getting to Know NLM Series. This series is designed to promote the different operational units at NLM, highlighting the major programs of each area and the skills, education, and expertise needed to succeed in each unit. Operational units covered in FY2000 were the Office of the Director, Office of Administration, Office of Communications and Public Liaison, and the Office of Equal Opportunity. The remaining operational units will be covered in FY2001. The series has been well received by NLM staff.
 - Collaborated with the Office of Communications and Public Liaison to promote the activities of the Diversity Council on the NLM Staff Bulletin Board located outside the cafeteria. This display has provided an excellent setting for celebrating the diversity found at the NLM.
 - Continued to coordinate scheduling of CPR training classes for NLM staff through the spring of 2000. To date approximately 45 NLM staff have received CPR training. The Council plans to continue scheduling CPR training classes in FY2001.
 - Spearheaded a food drive that resulted in NLM donating more than 18 boxes of nonperishable food items to the Shepherd’s Table, a community center for people in need.
 - Sponsored a highly successful employee appreciation event for all NLM federal and contract staff. A group photograph of all NLM employees was taken in front of the Library, followed by light refreshments and entertainment.

APPENDIX 1: REGIONAL MEDICAL LIBRARIES

1. **MIDDLE ATLANTIC REGION**
The New York Academy of Medicine
1216 Fifth Avenue
New York, NY 10029-5283
(212) 822-7396 FAX (212) 534-7042
States served: DE, NJ, NY, PA
URL: <http://www.nlm.nih.gov/mar>
2. **SOUTHEASTERN/ATLANTIC REGION**
University of Maryland at Baltimore
Health Science and Human Services
Library
601 Lombard Street
Baltimore, MD 21201-1583
(410) 706-2855 FAX (410) 706-0099
States served: AL, FL, GA, MD, MS, NC, SC, TN, VA, WV, DC, VI, PR
URL: <http://www.nlm.nih.gov/sar>
3. **GREATER MIDWEST REGION**
University of Illinois at Chicago
Library of the Health Sciences (M/C 763)
1750 West Polk Street
Chicago, IL 60612-7223
(312) 996-2464 FAX (312) 996-2226
States served: IA, IL, IN, KY, MI, MN, ND, OH, SD, WI
URL: <http://www.nlm.nih.gov/gmr>
4. **MIDCONTINENTAL REGION**
University of Nebraska Medical Center
Leon S. McGoogan Library of Medicine
Regional Medical Library
986706 Nebraska Medical Center
Omaha, NE 68198-6706
(402) 559-4326 FAX (402) 559-5482
States served: CO, KS, MO, NE, UT, WY
URL: <http://www.nlm.nih.gov/mr>
5. **SOUTH CENTRAL REGION**
Houston Academy of Medicine-
Texas Medical Center Library
1133 M.D. Anderson Boulevard
Houston, TX 77030-2809
(713) 799-7880 FAX (713) 790-7030
States served: AR, LA, NM, OK, TX
URL: <http://www.nlm.nih.gov/scr>
6. **PACIFIC NORTHWEST REGION**
University of Washington
Regional Medical Library, HSLIC
Box 357155
Seattle, WA 98195-7155
(206) 543-8262 FAX (206) 543-2469
States served: AK, ID, MT, OR, WA
URL: <http://www.nlm.nih.gov/pnr>
7. **PACIFIC SOUTHWEST REGION**
University of California, Los Angeles
Louise M. Darling Biomedical Library
Box 951798
Los Angeles, CA 90025-1798
(310) 825-1200 FAX (310) 825-5389
States served: AZ, CA, HI, NV and U.S. Territories in the Pacific Basin
URL: <http://www.nlm.nih.gov/psr>
8. **NEW ENGLAND REGION**
University of Connecticut Health Center
Lyman Maynard Stowe Library
263 Farmington Avenue
Farmington, CT 06030-5370
(860) 679-4500 FAX (860) 679-1305
States served: CT, MA, ME, NH, RI, VT
URL: <http://www.nlm.nih.gov/ner>

APPENDIX 2: BOARD OF REGENTS

The NLM Board of Regents meets three times a year to consider Library issues and make recommendations to the Secretary of Health and Human Services affecting the Library

Appointed Members:

FOSTER, Henry, M.D., Ph.D.
Senior Advisor to the President on Teen & Youth Issues
Department of Health and Human Services
Washington, D.C.

BARUCH, Jordan, Sc.D.
President, Jordan Baruch Associates
Washington, D.C.

BUNTING, Alison, M.L.S.
Associate University Library for Science
Louise Darling Biomedical Library
University of California, Los Angeles
Los Angeles, CA

KLEIN FEDYSHI, Michele, MSLS
Manager of Library Services
University of Pittsburgh Medical Center
Pittsburgh, PA

LEDERBERG, Joshua, Ph.D.
Sackler Foundation Scholar
Rockefeller University
New York, NY

LINSKER, Ralph, M.D.
IBM-T.J. Watson Research Center
Yorktown Heights, NY

NEWHOUSE, Joseph, Ph.D., Director
Division of Health Policy Research & Education
Harvard University
Boston, MA

PARDES, Herbert, M.D.
President and CEO
New York Presbyterian Hospital
New York, NY

PRIME, Eugenie, MS, MBA
Manager, Hewlett-Packard Libraries
Palo Alto, CA

WEICKER, Lowell, Governor
Greenwich, CT

Ex Officio Members:

Librarian of Congress

Surgeon General
Public Health Service

Surgeon General
Department of the Air Force

Surgeon General
Department of the Navy

Surgeon General
Department of the Army

Under Secretary for Health
Department of Veterans Affairs

Assistant Director for Biological Sciences
National Science Foundation

Director
National Agricultural Library

Dean
Uniformed Services University of the Health

APPENDIX 3: BOARD OF SCIENTIFIC COUNSELORS/ LISTER HILL CENTER

The Board of Scientific Counselors meets periodically to review and make recommendations on the Library's intramural research and development programs.

Members:

MARSHALL, Joanne G. Ph.D.
Dean, School of Information & Library Science
University of North Carolina
Chapel Hill, NC

MASYS, Daniel R., M.D.
Director of Biomedical Informatics
School of Medicine
University of California at San Diego
La Jolla, CA

MITRA, Sunanda, Ph.D.
Professor of Electrical Engineering
Texas Tech University
Lubbock, TX

SIEVERT, MaryEllen C., Ph.D.
Professor of Library and Information Science
University of Missouri
Columbia, MO

SRINIVASAN, Padmini, Ph.D.
School of Library & Information Science
University of Iowa
Iowa City, IA

APPENDIX 4: BOARD OF SCIENTIFIC COUNSELORS/ NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

The National Center for Biotechnology Information Board of Scientific Counselors meets periodically to review and make recommendations on the Library's biotechnology-related programs.

Members:

DELISI, Charles, Ph.D. (Chair)
Dean, College of Engineering
Boston University
Boston, MA

BLACK, Anne E., Ph.D.
Asst. Professor, Dept. of Physiology
Human and Molecular Genetic Center
Medical College of Wisconsin
Milwaukee, WI

LEE, Christopher J., Ph.D.
Assistant Professor
Molecular Biology Institute
University of California Los Angeles
Los Angeles, CA

MATISE, Tara Cox, Ph.D.
Department of Genetics
Rutgers University
Piscataway, NJ

PREUSS, Daphne K. Ph.D.
Assistant Professor
Molecular Genetics and Cell Biology
University of Chicago
Chicago, IL

APPENDIX 5: BIOMEDICAL LIBRARY REVIEW COMMITTEE

The Biomedical Library Review Committee meets three times a year to review applications for grants under the Medical Library Assistance Act.

Members:

BASLER, Thomas G., Ph.D. (Chair)
Chair, Department of Library Science &
Informatics
Medical University of South Carolina
Charleston, SC

ASH, Joan S., Ph.D.
Associate Professor
Library and Medical Informatics
Oregon Health Sciences University
Portland, OR

CHUEH, Henry C., M.D.
Co-Director, Laboratory of Computer Science
Assistant Professor of Medicine
Harvard Medical School
Boston, MA

CHUTE, Christopher G., Dr.P.H., M.D.
Section Head and Professor
Medical Informatics
Mayo Foundation
Rochester, MN

CLARKE, Neil D., Ph.D.
Associate Professor
Dept. of Biophysics and Biophysical Chemistry
Johns Hopkins School of Medicine
Baltimore, MD

DALRYMPLE, Prudence, Ph.D.
Dean and Associate Professor
Graduate School of Library Information Science
Dominican University
River Forest, IL

DIMITROFF, Alexandra, Ph.D.
Associate Professor
School of Library Science
University of Wisconsin
Milwaukee, WI

FUCHS, Rainer T., Ph.D.
Director, Research Informatics
Biogen, Inc.
Cambridge, MA

GUARD, J. Robert, MLS
Chief Information Officer
University of Cincinnati Medical Center
Cincinnati, OH

HUANG, H.K., D.Sc.
Director, Radiological Informatics
University of California at San Francisco
San Francisco, CA

MCGOWAN, Julie J., Ph.D.
Director, Ruth Lilly Medical Library
Indiana University School of Medicine
Indianapolis, IN

MILLER, Perry L., M.D.
Professor of Anesthesiology & Medical
Informatics
Yale School of Medicine
New Haven, CT

MILLER, Randolph A., M.D.
Chairman, Division of Biomedical Informatics
Vanderbilt University Medical Center
Nashville, TN

NILAND, Joyce C., Ph.D.
Chair, Division of Information Sciences
City of Hope National Medical Center
Duarte, CA

OHNO-MACHADO, Lucila, M.D., Ph.D.
Assistant Professor, Radiology Department
Brigham and Women's Hospital
Harvard Medical School
Boston, MA

ORTHNER, Helmuth , Ph.D.
Professor, Department of Health Informatics
University of Alabama
Birmingham, AL

PINSKY, Seth, Ph.D.
Senior Director
Merck and Company, Inc.
Rahway, NJ

SAHNI, Sartaj K., Ph.D.
Distinguished Professor
Computer & Information Science
University of Florida
Gainesville, FL

SHAVLIK, Jude W., Ph.D.
Professor of Medical Informatics
University of Wisconsin
Madison, WI

SWEENEY, Latanya K.
Assistant Professor of Computer Science
Carnegie Mellon University
Pittsburgh, PA

APPENDIX 6: LITERATURE SELECTION TECHNICAL REVIEW COMMITTEE

The Literature Selection Technical Review Committee meets three times a year to select journals for indexing in *Index Medicus* and MEDLINE.

Members:

COLLEN, Morris F., M.D.
Consultant and Director Emeritus
Kaiser Permanente Medical Care Program
Oakland, CA

BIRKMEYER, John D., M.D.
Assistant Professor of Surgery
Veterans Affairs Medical Center
White River Junction, VT

BOROVETZ, Harvey S., Ph.D.
Professor of Bioengineering
University of Pittsburgh School of Medicine
Pittsburgh, PA

COOPER, James N., M.D.
Director, INOVA Institute of Research
Chairman, Department of Medicine
Fairfax Hospital
Falls Church, VA

COPELAND, Robert L., Ph.D.
Associate Professor of Pharmacology
Howard University School of Medicine
Washington, D.C.

FUNK, Mark E.
Samuel J. Wood Library
Weill Medical College
Cornell University
New York, NY

LI, Yihong, Ph.D.
Assistant Professor
Oral Biology Department
University of Alabama School of Dentistry
Birmingham, AL

O'DONNELL, Anne Elizabeth, M.D.
Assistant Professor
Pulmonary and Critical Care Medicine
Georgetown University School of Medicine
Washington, D.C.

PICOT, Sandra J. Fulton, Ph.D.
Associate Professor
School of Nursing
University of Maryland
Baltimore, MD

TOLEDO-PEREYA, Luis H., M.D.
Director, Surgery Research & Molecular
Biology
Borgess Medical Center
Kalamazoo, MI

VALENTINE, Joan S., Ph.D.
Professor of Chemistry and Biochemistry
University of California
Los Angeles, CA

WILLIAMS, Benjamin T., M.D.
President
University Park Pathology Associates
Champaign, IL

APPENDIX 7: PUBMED CENTRAL NATIONAL ADVISORY COMMITTEE

The PubMed Central National Advisory Committee meets twice a year to review and make recommendations about the information resource, PubMed Central.

LEDERBERG, Joshua, Ph.D. (Chair)
Sackler Foundation Scholar
Rockefeller University
New York, NY

BROWN, Patrick O. Ph.D., M.D.
Associate Professor
Department of Biochemistry
Stanford University, School of Medicine
Stanford, CA 94305-5323

COZZARELLI, Nicholas, Ph.D.
Professor of Molecular and Cell Biology
Division of Biochemistry and Molecular
Biology
University of California
Berkeley, CA

DAVIDOFF, Frank, M.D.
Editor, *Annals of Internal Medicine*
Philadelphia, PA 19106

FRANCKE, Uta, M.D.
Professor of Genetics
Stanford University Medical Center
Stanford, CA

GINSPARG, Paul, Ph.D.
Theoretical Physicist
Los Alamos National Laboratory
Los Alamos, NM

HOMAN, Michael
Director of Libraries
Mayo Foundation
Rochester, MN

MARINCOLA, Elizabeth
Executive Director

American Society of Cell Biology
Bethesda, MD

McINERNEY, Suzanne
Health Writer/Patient Advocate
Hummelstown, PA

NEAL, James G.
Dean of University Librarians
Johns Hopkins University
Baltimore, MD 21218

RABB, Maurice F., M.D.
Professor of Ophthalmology
College of Medicine
University of Illinois at Chicago
Chicago, IL

ROBERTS, Richard J., Ph.D.
Research Director
Department of Bioinformatics
New England Biolabs
Beverly, MA

TRACZ, Vitek
Chairman and CEO
Current Science Group
Middlesex House
London, UK

VARMUS, Harold, M.D.
Director and CEO
Memorial Sloan-Kettering Cancer Center
New York, NY

WILLIAMS, James F., M.S.L.S.
Dean of Libraries
University of Colorado
Boulder, CO

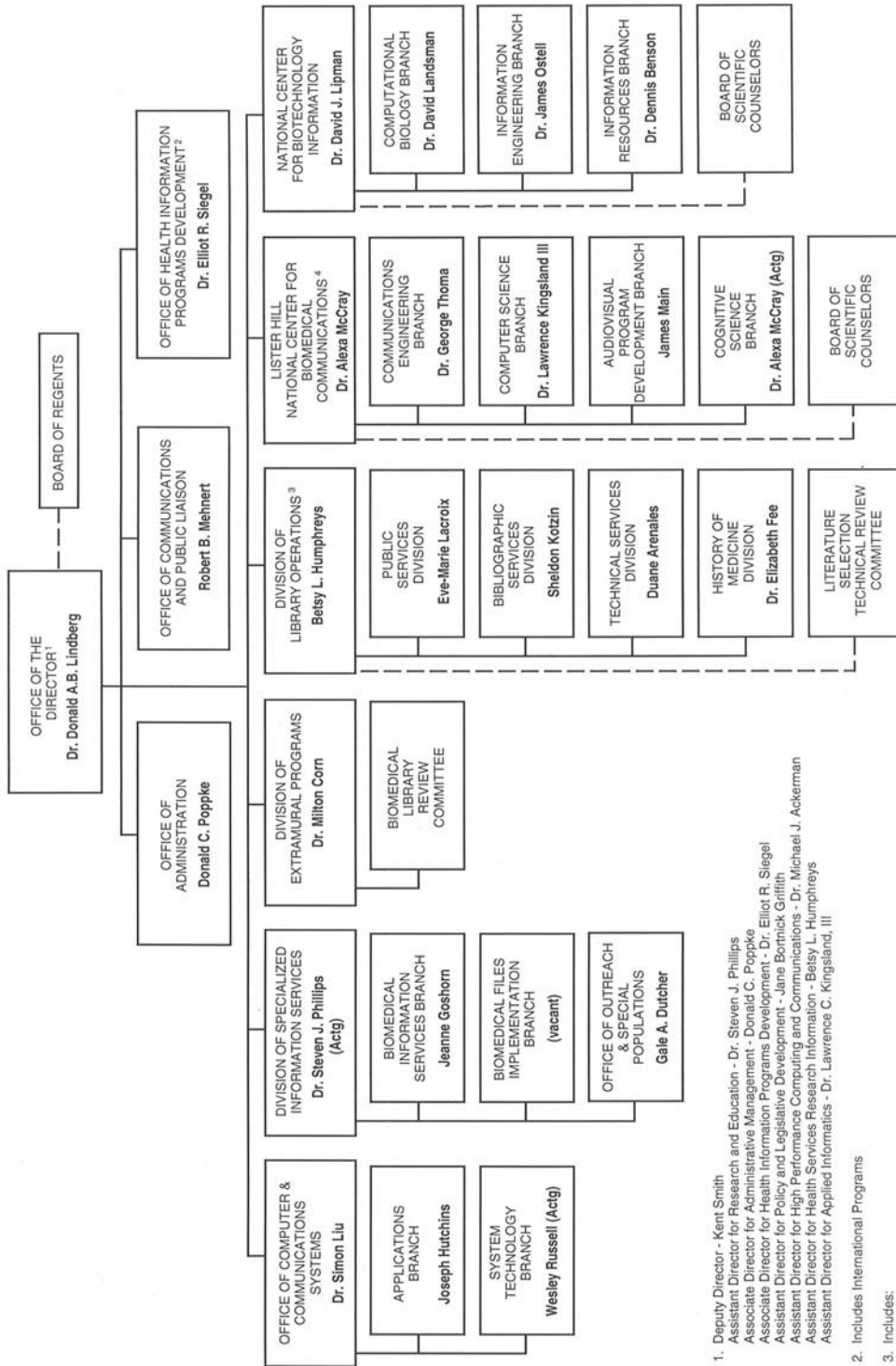
Further information about the programs described in this administrative report are available from:

*Office of Communications and Public Liaison
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894
(301) 496-6308*

E-mail: publicinfo@nlm.nih.gov

Cover: The names of the more than 700 employees who worked at the National Library of Medicine in Fiscal Year 2000.

National Library of Medicine



1. Deputy Director - Kent Smith
 Assistant Director for Research and Education - Dr. Steven J. Phillips
 Associate Director for Administrative Management - Donald C. Poppke
 Associate Director for Health Information Programs Development - Dr. Elliot R. Siegel
 Assistant Director for Policy and Legislative Development - Jane Bonnick Griffith
 Assistant Director for High Performance Computing and Communications - Dr. Michael J. Ackerman
 Assistant Director for Health Services Research Information - Betsy L. Humphreys
 Assistant Director for Applied Informatics - Dr. Lawrence C. Kingsland, III

2. Includes International Programs

3. Includes:
 National Network of Libraries of Medicine, Head - Dr. Angela Ruffin
 Medical Subject Headings Section - Stuart Nelson, M.D.
 National Information Center on Health Services Research and Health Care Technology - Marjorie A. Cahn

4. Office of High Performance Computing and Communications, Head - Dr. Michael J. Ackerman