Office of Health Information Programs Development

Elliot R. Siegel, Ph.D.
Associate Director

International Programs

The year 1998 saw continuing leadership by the National Library of Medicine in several areas of international work: global Internet connectivity, Internet connectivity and access to information for malaria researchers in Africa, and the Board of Regents' approval of a Long Range Planning Panel on International Programs Report followed by planning of the first meeting for its implementation.

Internet communications connectivity and access to information resources by malaria research scientists in Africa is part of the Multilateral Initiative on Malaria, a multilateral initiative undertaken in collaboration with the National Institute of Allergy and Infectious Diseases, the Fogarty International Center, and the Office of the Director, National Institutes of Health.

Future directions of NLM's international programs received substantial attention as a new long-range plan was approved by the NLM Board of Regents and plans for implementation began. Other international activities were carried out with individual countries and governmental and nongovernmental organizations. Information management training was provided to colleagues from abroad, numerous professional visitors were received from around the world, and publications were exchanged with libraries in other countries.

Long Range Planning Panel on International Programs

The NLM Board of Regents approved the NLM Long Range Planning Panel on International Programs Report in 1998. This panel was chartered by the NLM Board of Regents in May 1995 to advise NLM on the relative priority of its international activities and responsibilities and to assist in the development of appropriate strategies consistent with the Library's statutory mission and available resources. The Panel was chaired by former NIH Director Donald S. Fredrickson and consisted of 25 distinguished members assisted by expert consultants in the fields of medicine, telecommunications, health sciences librarianship, electronic publishing, and related fields. The Panel, which met three times in 1996-1997, made the following formal recommendations to the Board on the future of NLM's international activities:

- Objective 1: Strengthen and expand global access to the world's health-related literature
- Objective 2: Chart new routes to biomedical knowledge and its use
- Objective 3: Enable NLM to fulfill its international mission

The Panel's report was published in the fall of 1998. NLM will be working to identify ways to implement the plan through collaborative partnerships with institutions, organizations, and other entities around the world.

Internet Connectivity at Malaria Research Sites in Africa

> *What a pleasure for us and our collaborators to sit in our offices and browse the Web sites, being in contact with the world in a few seconds, looking for the hidden world. What a great potential we are discovering.*

This testament of Dr. Yeya Toure Director of the Malaria Research and Training Center in Bamako, Mali, reveals at once the excitement of African scientists and the potential that Internet access holds for scientific research. Since June 1998 he and his colleagues at the MRTC have had access to the Internet and the World Wide Web through microwave technology. The equipment (including a local area network on site), installation, and training were funded by the National Institute of Allergy and Infectious Diseases, the NIH Office of the Director, the National Library of Medicine, and the World Bank.

This story from Mali is the first chapter in the work of the Multilateral Initiative on Malaria Communications Working Group (MIM CWG), chaired by NLM. Featured prominently are the major MIM objectives—support for African scientists, the ability of malaria researchers to connect with one another and sources of information, 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.

In addition to the malaria research site in Bamako, Mali, the MIM/CWG endorsed five more locations for the initial connectivity phase. They are: in Kenya, the Centers for Disease Control/Kenya Medical Research Institute (KEMRI) site in Kisian and the Wellcome Trust/KEMRI site in Kilifi, and in Tanzania, three sites of the National Institute of Medical Research in Dar es Salaam, Ifakara, and Amani. In several instances, sites have been provided with computer equipment and training, but since the majority of them are in relatively remote locations, traditional means of connecting to the Internet are not viable due to unreliability or bandwidth restrictions.

Subsequently, NLM supported site visits and assessments, consultancies, evaluation and testing of the extant technology. Related issues of user training, in-country licensure of technology, and allowances for future technological advances (such as predicted worldwide availability of low-cost commercial satellite systems) all figured in the development of a draft implementation plan by the MIM/CWG.

The plan 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 cooperative research and disseminate their results. Recommended technologies are VSAT, which uses a geostationary satellite and a small earth station, and microwave, which uses radio waves. The latter is less expensive but is limited to line of sight transmission. The MIM sites that wish to operate a radio or VSAT link will have to gain permission from the relevant in-country authority.

With the Mali model fully operational, NLM/NIH has stepped forward with an offer to fund the upfront equipment purchase and installation costs at these five sites, if partner funders can commit to support ongoing operational costs. Sustainability is an essential ingredient if lasting connectivity is to be achieved for these research sites. In the case of the CDC/KEMRI site in Kisian, Kenya, a funding partnership between the NLM and the CDC is firmly in place.

Global Internet Connectivity

In 1998, NLM continued and expanded its end-to-end Internet connectivity testing and evaluation project. This project is intended to explore the methods and metrics needed to better understand the quality of Internet performance from the end user perspective. To do this, NLM is using test methods that measure, for example, the size of the Internet transmission "pipe," the round-trip-time for sending packets of information to the destination and back, packet "loss," and response time. NLM is collaborating with numerous domestic and international partners to test Internet pathways around the world.

Currently, NLM monitors about 85 network paths to hosts (typically web servers) in 29 countries and in 14 time zones. Hosts are located in all major regions of the world. The monitoring includes locations in all G-7/8 countries. NLM is conducting symmetrical (two-way) and time zone testing with three partners as part of the G-7/8 Internet connectivity initiative: the Western Universities Research Consortium at the University of Calgary, Canada; OMNI (Online Medical Networked Information) at the University of Nottingham, England; and ForthNet, SA, in Crete, Greece. In addition, NLM has conducted special testing of Internet pathways to sites in several sub-Saharan African countries and in Mexico and Russia.

Results of Phase I of the Internet connectivity project were published by Fred B. Wood, Victor H. Cid, and Elliot R. Siegel, in "Evaluating Internet End-to-End Performance: Overview of Test Methodology and Results," *Journal of the American Medical Informatics Association*, Vol. 5, November/December 1998, pp. 528-545.

Activities With International and Multinational Organizations

Working in collaboration with the International Council for Scientific and Technical Information, NLM helped to successfully conclude a pilot demonstration project to reduce existing technological and tariff barriers to the flow of scientific and health-related information over the Internet to developing regions of the world. Focused in the eastern Caribbean, this effort was carried out in conjunction with the United Nations Educational, Scientific, and Cultural Organization (UNESCO), the Pan American Health Organization (PAHO), the International Telecommunications Union, and the United Nations Development Programmes.

In January 1997, NLM hosted the first U.S. meeting of the national coordinators of the G-7 Global Healthcare Applications Project. At this meeting, the coordinators approved a U.S. proposal for global Internet connectivity advanced by NLM. The premise of this proposal was that many G-7 Global Healthcare Applications projects depend in whole or in part on the availability of the Internet. As the preferred technical means for the exchange of biomedical information with and among the G-7 nations and beyond, the Internet is a critical component of the emerging global health information infrastructure. NLM has begun to lay the groundwork for collaborative Internet testing that should improve understanding of Internet connectivity from the user's perspective and ultimately help to better inform users about selecting Internet services and

solving problems in use of the Internet. NLM has identified a need for objective and widely accepted methods of characterizing and diagnosing the quality and performance of Internet connections under different conditions (e.g., locations, distances, time periods, and data volumes). Also, NLM has established collaborative relationships with an initial group of domestic and international partners both to validate testing methods and to conduct tests of Internet connectivity between selected locations.

A second U.S.-led project, the Multi-Language Anatomical Digital Database, which was developed by NLM, was also approved by the coordinators of the G-7 Global Healthcare Applications Project. This project takes advantage of the multilingual capabilities of the Unified Medical Language System and facilitates multilingual access to the future Visible Human Database. This database consists of images that are inherently word free and language neutral; thus, with the addition of multilingual anatomical labeling, these images can serve multilingual needs. The project will serve as a test bed for a range of applications, including medical education, patient information, and telemedicine consultation. The multilingual capability is especially advantageous for international applications.

International Visitors

NLM continues to be a focal point for visitors of the international community from a variety of disciplines. Many of these visitors are responsible for medical, scientific, or technical information in their own countries. Visitors are officially received and briefed on relevant aspects of NLM operations and research. Among the visitors in 1997 were representatives from:

Albania, Armenia, Azerbaijan,
Bolivia, Bosnia, Cameroon, China,
Czech Republic, Egypt, Georgia,
Germany, Haiti, India, Indonesia,
Israel, Jamaica, Japan, Kazakhstan,
Kenya, Laos, Macedonia, Mali,
Mexico, Mongolia, Mozambique,
Nigeria, Norway, Peru, Poland,
Russia, Senegal, South Africa,
Sweden, Senegal, Switzerland,
Taiwan, Tanzania, Thailand,
Uganda, United Kingdom,
Yugoslavia, Zambia, Zimbabwe

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 heath 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 re view 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 ability to conduct this 5-year review was, to a great extent, facilitated by data collecting and reporting by the Regional Medical Libraries and NN/LM. This is an important first step in assessing the strategies that have been undertaken so far. However, to evaluate outreach approaches better and demonstrate the effectiveness of outreach projects in the future, several additional steps should be taken. NLM should develop benchmarks against which the RMLs and the NN/LM can measure the effectiveness of their outreach efforts. In addition, NLM and the RMLs should work together to develop further expertise in evaluation methodology. Evaluation components should be an integral part of all NLM-sponsored outreach.

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.

For this project, with its focus on outreach evaluation, NLM selected the NN/LM's Pacific Northwest Regional Medical Library, located at the University of Washington Health Sciences Library. NLM decided that a sub-focus on Native Americans in the Pacific Northwest would be useful to better understand outreach to underserved populations and test the outreach planning and evaluation model to be developed by the project. Initial project plans were presented at an outreach panel as part of the 1997 Medical Library Association meeting in Seattle. The project has established an advisory panel consisting of representatives from academia, libraries, health professionals, and Native American groups. Finally, a separate but related tribal connectivity project also to be administered by the Pacific Northwest RML will provide support for new or enhanced Internet connections for selected American Indian tribes and Alaska Native villages in the Pacific Northwest. Some of these sites will be used for pilot tests of the outreach planning and evaluation guide.

Library Operations

Lois Ann Colaianni
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 provides primary health sciences library services throughout the country. LO's basic information services provide an essential foundation for NLM's Outreach Program and for its biotechnology, AIDS, and health services research information programs.

The largest of NLM's Divisions, LO has a multi-disciplinary staff of librarians, technical information specialists, subject experts, health professionals, historians, and technical and administrative support personnel. This staff also directs the National Information Center on Health Services Research and Health Care Technology (NICHSR); plays a key role in the development of NLM's Web information services; carries out an active program of research in the history of medicine; conducts research and evaluation studies related to the Library's programs and services; directs and participates in research in advanced information storage and retrieval; and contributes to the development of Federal health data policy. LO staff members participate actively in Library-wide efforts to improve the quality of work life at NLM, including the Diversity Council and the expansion of the NLM Intranet.

Planning and Management

In FY 1998, LO's major priorities were to respond effectively to the growing number and changing nature of user demands generated by free MEDLINE access and to make steady progress on replacing the legacy automated systems that support internal operations and library services. As expected, the availability of free MEDLINE increased the general public's use of NLM's Web-based services and led to requests from the public for documents and reference assistance. While developing interim approaches to responding to these requests, LO also participated in NLM-wide planning for new and enhanced services designed to meet the needs of these users.

LO continued to work with other NLM Divisions to manage the complex technical, resource, and scheduling issues related to the Library's System Reinvention effort. Several major System Reinvention goals were met during the year, including full implementation of the Relais document delivery processing system, provision of Web-based access to all of NLM's databases, and significant progress toward implementation of a commercial integrated library system (ILS).

The ongoing transition away from the ELHILL mainframe-based retrieval system raised concerns about whether NLM's new Web-based retrieval systems would support the advanced searches frequently performed by expert searchers and whether all NN/LM librarians would have the training and connections necessary to use the new systems effectively before ELHILL access was discontinued. In April 1998, NLM and the Medical Library Association convened a meeting of health sciences librarians to review problems being encountered in accessing and using the Web-based systems and to discuss PubMed and Internet Grateful Med enhancements useful to expert searchers, strategies for providing additional training opportunities, and ways for NLM to communicate upcoming changes to librarians more effectively.

Early in the fiscal year, NLM selected the Voyager ILS to replace the custom mainframe-based legacy systems currently used for acquisitions, serials processing, cataloging, and onsite circulation. Substantial staff resources were then devoted to specifying, developing, and testing the programs that will convert data from more than 15 existing files to the ILS; to pre-conversion data clean-up; to specifying and testing Voyager enhancements needed for NLM functions; and to extensive staff training in preparation for initial implementation in early FY 1999. As expected, the reallocation of staff to ILS implementation tasks caused small decreases in production levels for some acquisitions and cataloging tasks.

LO continued the consolidation of customer service operations by transferring the primary responsibility for responding to online user inquiries from the MEDLARS Management Section (MMS) to the Reference Section. In turn, MMS assumed lead responsibility for expanding the number of publishers supplying SGML-tagged citation data for MEDLINE from the National Center for Biotechnology Information (NCBI) and for contracting out regular distribution of MEDLINE data via ftp from the Office of Computer and Communications Systems (OCCS).

Additional information about the development of pilot services for the general public and specific system reinvention activities appears elsewhere in this and other chapters.

Collection Development and Management

Many NLM services depend on the Library's comprehensive collection of biomedical literature. LO ensures that the NLM collection meets the needs of current and future health professionals and researchers by developing and updating a formal literature selection policy, acquiring and processing literature that meets its selection guidelines in all languages and formats, organizing and maintaining the collection for efficient current use, and preserving materials acquired for future generations. At the end of FY 1998, the NLM collection included 2.26 million volumes and 3 million other items, including audiovisuals, microforms, computer software, pictures, and manuscripts. (Table 2)

Selection

LO and its book dealers and subscription agents select literature for the NLM collection according to the guidelines in the *Collection Development Manual of the National Library of Medicine*, which has normally undergone a major review and revision every 5 to 7 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 FY 1998, LO staff began routine selection of some materials published only in electronic format.

LO also reviews segments of the collection to determining the extent to which it has been successful in adhering to selection guidelines for particular subject fields. In FY 1998, LO staff examined the Library's collection of anthropology serials and canceled some subscriptions as a result. A small collection assessment study of health policy literature led to expanded efforts to obtain this material, with an emphasis on acquisition of technical reports and other "gray" literature.

Acquisitions

LO received and processed 168,563 contemporary books, serial issues, audiovisuals, and software packages. (Table 3) A net total of 37,734 volumes and 32,133 other items (e.g., microforms, audiovisuals, software, manuscripts, pictures) were added to the NLM collection in FY 1998. Blanket purchase agreements for monographs were extended through FY 1999, and staff began use of credit cards for selected literature purchases to reduce acquisition delays. Acquisitions and serials processing staff devoted substantial effort to developing specifications for converting records from NLM's legacy systems to the new ILS and to completing data clean-up tasks that would allow more accurate record conversion.

Among the important historical books and manuscripts acquired by the Library this year were Bartholomaeus Montagnana's *Consilia Medica* (Padua, 1476), the first known set of printed clinical records; Jean Fernel's *La Chirurgie Translatee de Latin en Francois* (Sens, 1579), a rare first edition; the *Zene Artzney* (Frankfurt 1541), the first known printed work devoted entirely to dental care; Philipp Ulstadius' *Coelum Philosophorum* (Strasburg, 1536), an early German work on distilling theory and practice; Eucharius Rosslin's *Kreuterbuch* (Frankfurt, 1550), the most popular herbal of all time; *Commentaires Tres Excellens de l'Hystoire des Plantes* (Paris, 1549), the first edition in French of Leonhart Fuchs' great herbal *De Historia Stirpum;* Nachala Obshchei Voennopolevoi Khirurgii (Dresden, 1865-1866), the first edition in Russian of Nicolai Pirogov's major treatise on military surgery; and Guia de parteiras por perguntas e respostas, an unpublished manual of obstetrical knowledge designed for midwives.

The Information Office of the Director of NIH donated 12 motion pictures and the Medical Arts Department donated more than 2,200 posters to the historical collections; Mrs. Johanna Kluger donated an oil portrait of Dr. Albert Schweitzer by Eric Voight; and Mr. William continued his generosity to NLM by donating a substantial collection of posters, still images, and ephemera.

Three NIH scientists who received the Nobel Prize in Medicine: Drs. Julius Axelrod, Christian Anfinsen, and Martin Rodbell, agreed to donate manuscript collections to NLM. The Library has received selected manuscripts from Nobel Laureate Dr. Joshua Lederberg of Rockefeller University. Former NIH Director Dr. Donald S. Fredrickson donated additional manuscripts to his collection of papers already at NLM. Taken as a whole, these valuable collections provide unique documentation on the history of post-World War II science, and more specifically on the history of biomedical research and discovery at the National Institutes of Health. These collections form the core of the Library's newly established Digital Archives Program and will be organized, cataloged, and digitized for access on the World Wide Web.

Preservation and Collection Management

To extend the life of NLM's archival collection and to keep it accessible and in good order, LO carries out a range of preservation and collection management functions. These functions include:

binding, microfilming, conserving rare and unique items, 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 also distributes data in MARC format about what NLM has preserved and provides preservation information useful to other health sciences libraries on NLM's Web site. NLM continues to monitor developments in preservation technology and to promote the use of more permanent media in new biomedical publications.

In FY 1998, LO bound 24,323 volumes, microfilmed 5,503 volumes, conserved 231 historical volumes and audiovisuals, and repaired 2,571 volumes in the onsite book repair and conservation laboratory. Collection management staff resolved or referred for repair 11,802 items with bibliographic or physical problems. The current binding contract was extended for a year, pending availability of the ILS binding module, and the microfilming contract was modified to include some color and continuous tone filming. The oil portrait of Fielding Garrison, a major figure in the development of NLM and historical medical bibliography, was restored and hung in the Incunabula Room in the History of Medicine Division. All oil paintings in the NLM collection which are not on display in the Library were rehoused in new storage bins designed especially for them. Following the discovery that some items had apparently been taken many years ago, a contract was awarded to inventory the 1800-1914 printed works collection.

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. LO also maintains the Medical Subject Headings (MeSH), the thesaurus used by NLM and many other institutions to describe the subject content of biomedical information; links MeSH to other biomedical vocabularies within the Unified Medical Language System (UMLS) Metathesaurus; and produces the *National Library of Medicine Classification*, a scheme for arranging physical library materials according to their subject matter used by health sciences libraries around the world.

Thesaurus Development

The 1999 MeSH contains 19,232 main subject headings, 787 subheadings or qualifiers, 122 publication types, and more than 103,500 supplementary records for chemicals and other substances. Changes made for the 1999 edition include 303 new descriptors, updated names for 49 main headings, and 507 new cross-references. Special efforts were undertaken to eliminate the backlog of requests for new MeSH headings, to enhance and update the terminology for fungi, to reorganize the hierarchical arrangement of hematologic disorders and vascular diseases, and to develop a pharmacologic action tree. The number of publication types was expanded to accommodate the Cataloging Section's plan to eliminate the use of form subheadings in catalog records and to rely on the publication type to supply comparable information. LO staff worked with the Office of Computer and Communications System on the development of a new Oracle-based MeSH data creation and maintenance system to be implemented in 1999.

The majority of the editing for the 1999 edition of the UMLS Metathesaurus was completed in FY 1998 under MeSH Section supervision. New additions to the 1999 Metathesaurus will include: the Beth Israel Clinical problem list vocabulary; the *Alcohol and Other Drug Thesaurus*; clinical drug terminology derived from *Micromedex*; the *Pharmacy Practice Activity Classification*; the *Patient Care Data Set*, which contains detailed nursing terminology; *Alternative Billing Concepts*, used to bill for procedures by licensed practitioners of alternative therapies; a small initial set of valid values for segments of HL7 messages; and terminology used by the National Cancer Institute to characterize cancer research projects. The coverage of non-English terminology will be expanded to include the German translations of ICD-10 and the *Universal Medical Device Nomenclature System*, the French, German, Portuguese, and Spanish translations of the *International Classification of Primary Care*, and the Italian translation of portions of MeSH.

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 has begun to catalog information sources that are available via the Web, but not physically held in the NLM collection. Many health sciences libraries also use MeSH and the NLM Classification to catalog materials of local interest for which no NLM catalog record is produced.

In FY 1998, the Technical Services Division cataloged 18,803 contemporary books, serials, nonprint items (including some Web-based information sources), and cataloging-in-publication

galleys, using a combination of in-house staff and contractors. The working inventory of uncataloged books held steady at about 4 months workload of primarily non-English items. The Cataloging Section completed a major revision in subject cataloging policy to adopt the indexing form of subject headings effective in 1999. This should facilitate integrated searching of books and journal articles as NLM's system reinvention progresses. Cataloging staff were also heavily involved in data cleanup, data conversion, and staff training in preparation for the implementation of the new Integrated Library System.

The History of Medicine Division cataloged 193 early books. LO staff worked with the Lister Hill Center to establish a production operation for scanning and creating metadata for the manuscripts of Nobel prize-winning scientists, with an emphasis on those who have worked at NIH. During FY 1998 work focused on the papers of Joshua Lederberg. The portion of these papers dealing with Oswald T. Avery became the first collection accessible from NLM's new *Profiles in Science* Web site which debuted in September 1998.

Indexing

LO indexes articles from nearly 4,000 biomedical journals so that users of MEDLINE and the products generated from it can locate articles on specific biomedical topics. A combination of in-house staff, contractors, and cooperating U.S. and international organizations perform the indexing under the supervision of the Index Section. 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 (Appendix 6), an NIH-chartered committee of outside experts, advises NLM about which journals should be indexed for MEDLINE, *Index Medicus*, and other NLM databases. In FY 1998, the LSTRC reviewed 349 journal titles and rated 64 sufficiently highly for immediate inclusion in MEDLINE. The Committee also reviewed recommendations on journal coverage from professional societies in the fields of anesthesiology and critical care medicine. As a result of these reviews, the LSTRC recommended that NLM discontinue indexing 5 titles. Twelve consumer health publications were added to MEDLINE in January as a result of a special review by experts in consumer health information services from across the country. In collaboration with the NIH Office of Alternative Medicine, LO arranged for a review of more than 650 alternative medicine journals by 16 U.S. and international alternative medicine research centers. The purpose of the review was to determine whether there were additional alternative medicine journals suitable for indexing in MEDLINE or addition to the NLM collection. As a result of this review, 7 journals were added to MEDLINE.

In FY 1998, NLM added 411,921 indexed citations to MEDLINE, 21 percent less than in FY 1997 when LO was clearing a substantial backlog caused by the suspension of the data entry contract in FY 1996. To avoid similar disruptions in the future, NLM now has three parallel input streams for MEDLINE citations and abstracts: traditional keyboarding, scanning and optical character recognition (OCR), and direct electronic input from publishers. In FY 1998, 60 percent of MEDLINE citations and abstracts were keyboarded, 30 percent were entered via scanning and keyboarding, and 10 percent were received in tagged electronic format from publishers. By the middle of the year, the Library had achieved its goal of one-third of current MEDLINE input via scanning and OCR, using a system developed and then incrementally enhanced by the Lister Hill Center. The OCR system uses a spell-checking dictionary derived from the UMLS Metathesaurus and other sources to reduce the number of correctly spelled words that are flagged as OCR errors. At the end of FY 1998, 85 different publishers were providing electronic citation and abstract data to NLM for a total of 382 journals. LO is actively seeking electronic data from additional publishers and the amount of data received via this mechanism is growing steadily.

Regardless of the initial data entry method, all citations and abstracts are transferred to the online indexing system where indexers add subject headings and other data elements necessary to complete MEDLINE records. The online indexing system is one of many legacy systems that will be replaced as part of NLM System Reinvention. During FY 1998, LO staff worked with the Office of Computer and Communications to define the functional requirements for the reinvented indexing system and the ways that it will interface with the PubMed retrieval system and the Integrated Library System.

Information Products

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

Databases and Web Information Resources

Use of MEDLINE and other NLM databases continued to increase dramatically to an estimated annual total of 104,000,000 searches. (Table 7) Of these, only four million were direct searches of ELHILL, the mainframe retrieval system that is being gradually phased out as part of NLM System Reinvention. As of September 1998, all NLM databases are available free via the Web. Web users may search MEDLINE in PubMed, use the Internet Grateful Med interface to search MEDLINE and 13 other databases which are currently still available in ELHILL, and access the Specialized Information Services toxicological and environmental health databases on the new TOXNET Web server. (See Chapter 2 for additional information about TOXNET). Now that all NLM databases are available free-of-charge via the Web, LO has phased out flat-rate and fixed-fee arrangements for access to ELHILL databases.

During FY 1998, LO assisted the National Center for Biotechnology Information in the design and testing of three significant upgrades to PubMed, principally to improve advanced search capabilities. LO also collaborated with the Lister Hill Center to establish procedures for providing updated UMLS/MeSH mapping tables to NCBI for use in PubMed. LO staff also assisted the Lister Hill Center with the design and testing of a new version of Internet Grateful Med which uses PubMed as its underlying search engine for MEDLINE and ELHILL as the search engine for 13 other NLM databases. To help expert searchers in making the transition from ELHILL to the new Web-based retrieval systems, LO worked with the Online Training Center at the New York Academy of Medicine to develop a new online training program which emphasized PubMed/Internet Grateful Med searching for experienced searchers. The number of classes offered was expanded significantly to offer additional training opportunities to interested health sciences librarians; 1,884 people attended online training classes in FY 1998, an increase of 225 percent from the previous year. Some PubMed and Internet Grateful Med training materials were also made available via the Web. NLM ceased distribution and support of Grateful Med disks for DOS and Windows.

Citations from the 1960-63 volumes of *Index Medicus* were converted to machine-readable form and will become available online in OLDMEDLINE in early FY 1999. NLM will continue the gradual conversion of earlier years of *Index Medicus* for addition to OLDMEDLINE. LO worked with the Lister Hill Center to expand HSTAT (Health Services and Technology Assessment Text) to incorporate additional full-text documents produced by the Agency for Health Care Policy and Research, other components of NIH, and the Substance Abuse and Mental Health Administration. Links were established between PubMed and HSTAT for documents that are indexed in MEDLINE and available in full-text in HSTAT. About 18,000 historical images were rescanned to improve the quality of the pictures available via the Web in *Images from the History of Medicine*.

During FY 1998, LO worked with OCCS to develop a pilot consumer health Web information resource to be released in early FY 1999. This experimental service, called MEDLINEplus, will provide tailored consumer-friendly access to MEDLINE; targeted subject access to high quality consumer health and patient information produced by NIH, other HHS agencies, and other authoritative sources; and links to Healthfinder, self-help groups, clearinghouses, health-related organizations, and clinical trials. MEDLINE*plus* relies on an innovative combination of relational database, Cold Fusion, Java scripts, and Web technology to allow efficient and distributed creation, review, and maintenance of links to consumer health information by librarians working in widely dispersed geographic locations. LO plans to apply this same technology to other NLM Web pages that include multiple links to outside resources, including the expanded directory pages for health services research and public health information sources produced by LO's National Information Center on Health Services Research and Technology Assessment (NICHSR).

Machine-Readable Data

NLM leases its databases in machine-readable form to promote the broadest possible use of its authoritative 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 or use them to improve the functionality of a variety of medical information systems. In FY 1998, NLM distributed MEDLINE and other bibliographic databases to 57 different organizations. Hundreds of organizations and individuals obtained MeSH data via file transfer protocol (ftp) from the Internet. The UMLS Knowledge Sources and associated lexical programs were sent to more than 800 licensees world-wide; licensed users also have access to the UMLS data on the Internet-based UMLS Knowledge Source Server, developed and maintained by the Lister Hill Center.

In FY 1998, NLM updated its MARC distribution format for bibliographic records to the March 1997 version of MARC. The Library conducted a successful experiment with third-party distribution of MEDLINE data via ftp and DAT tape

and intends to expand this into an operational service. LO initiated quarterly distribution of MeSH supplementary concept records (for chemicals, other substances, protocols, etc.) via ftp from the MeSH Section Web pages.

Print and Electronic Publications

NLM publishes some of its authoritative data in printed publications, including *Index Medicus* and several MeSH publications. The Library continues to review and modify or eliminate specific publications that have outlived their usefulness, given increasing user access to the more flexible electronic forms of NLM data. In FY 1998, NLM ceased publication of the monthly *Abridged Index Medicus*.

The NLM World Wide Web site continues to grow as the primary vehicle for distributing a wide range of NLM publications B from Fact Sheets to extensive reports. In FY 1998, LO converted the *Technical Bulletin* to a Web publication and discontinued the printed version. The Web format allows immediate publication of individual articles and links to additional supporting information on NLM's Web site. The Web is an excellent publication medium for a new *Catalogue Raisonnee* of NLM's Persian and Arabic Manuscript collection that is currently under development. Many of the manuscripts have beautiful color illuminations and illustrations that would be prohibitively expensive to include in a printed publication. Users at an average of 271,331 distinct Internet hosts each quarter accessed copies of publications on NLM's Web site 1,548,966 times in FY 1998.

LO produces a series of *Current Bibliographies in Medicine*, which are available on the Web site and, in some cases, also in print. Each bibliography addresses a topic of current interest that may be difficult to search in NLM's databases or be spread across the literature of multiple disciplines. The bibliographies cover subjects important to specific programs of NLM, NIH, or other Federal agencies and are often produced in conjunction with NIH Consensus Development Conferences or other special meetings. LO staff members collaborate with outside experts to produce each bibliography. FY 1998 additions to the series included: *Zinc and Health*, *Diagnosis and Treatment of Attention Deficit Hyperactivity Disorder*, *Rehabilitation of Persons with Traumatic Brain Injury*, and *Effective Medical Treatment of Heroin Addiction.*

User Services

In addition to its electronic and printed products, LO provides document delivery, reference, and customer service as a national back-up to services available from other health sciences libraries and information suppliers; coordinates the National Network of Libraries of Medicine (NN/LM) which works to equalize access to information services and technology throughout the United States; arranges exhibitions and programs on the history of medicine and related fields; and conducts training programs for health sciences librarians.

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 supplied from other NN/LM libraries or document suppliers. LO also retrieves documents from NLM's closed stacks for use by onsite patrons. In FY 1998, the Collection Access Section received a total of 694,281 requests for documents, a 10 percent increase from the previous year. (Table 6) Requests from other libraries increased 6 percent; requests from onsite users increased 15 percent. NLM filled 74 percent of the interlibrary loan requests it received and processed 90 percent within 24 hours of receipt.

In FY 1998, LO implemented Relais, an automated document delivery system originally developed for the Canada Institute of Scientific and Technical Information (CISTI) and modified by its creator (Network Support, Inc.) to meet NLM's requirements. A part of NLM System Reinvention, Relais captures incoming DOCLINE requests directed to NLM and document requests generated in NLM's Reading Rooms, sorts and prints the requests for processing, and allows LO staff to track their whereabouts in the processing stream. Documents are scanned, rather than photocopied, and the temporarily stored images are sent either to a high-speed printer or directly via Internet or fax, depending on the requester's preference. In FY 1998, 19 percent of NLM's filled interlibrary loan requests were delivered via Internet or fax.

Relais enhances the processing of document requests onsite at NLM, but the DOCLINE system continues to route ILL requests among more U.S. and some international health sciences libraries. The 3,023 active DOCLINE users entered 2.9 million DOCLINE requests in FY 1998, an increase of 1 percent from FY 1997. As in past years, 94 percent of the requests were filled. Use of the Loansome Doc feature, which allows individuals searching via PubMed or Internet Grateful Med to route document requests through a health sciences library which agrees to serve them, increased 29 percent in FY 1998; more than 400,000 documents were requested via this mechanism.

Many international PubMed and Internet Grateful Med searchers are requesting Loansome Doc service. In an effort to serve these users, LO has

enlisted a number of international libraries which are willing to supply documents to individuals outside the U.S. and has provided information about these libraries on the Loansome Doc registration page on NLM's Web site. At the end of FY 1998, the following libraries were identified as international Loansome Doc suppliers: Centro Latino Americano y del Caribe de Informacion en Ciencias de la Salud (BIREME) in Brazil; Canada Institute of Scientific and Technical Information (CISTI); the National Library of Health Sciences in Finland; the Chinese University of Hong Kong; Centro National de Informacion y Documentacion sobre Salud (CENIDS) in Mexico; the State Central Scientific Medical Library in Russia; and Deutsche Zentralbibliothek fur Medizin (ZBMed) in Germany. NLM also provides information about U.S. health sciences libraries willing to provide document delivery services to libraries anywhere in the world.

Reference and Customer Service

LO provides reference and research assistance to onsite and remote users as a backup to services available from other health sciences libraries. The LO staff also responds to inquiries from the Library's users who are seeking information about NLM products and services or assistance in using these services. In FY 1998, the Reference Section and the History of Medicine Division received a total of 83,239 reference inquiries, up 14 percent from FY 1997. The combined total of e-mail inquiries received by the Reference Section and online services inquiries received by the MEDLARS Management Section between November 1, 1997 and September 30, 1998 was 26,804. Use of e-mail to submit inquiries increased substantially from the previous year. Mail from a variety of NLM e-mail addresses was gradually redirected to a single initial customer service entry point during this period to be tracked in the CustomerQ software. Plans are proceeding to record all phone inquiries in CustomerQ in FY 1999. This will allow more comprehensive analysis of questions received and help NLM to assess the impact of its new consumer health information services. The analysis of e-mail inquiries recorded in FY 1998 has already prompted the creation of "Frequently Asked Questions" (FAQs) pages on NLM's Website and a revised form for submitting inquiries to the Website.

During FY 1998, NLM's services to onsite users were enhanced by the development of special Reading Room Web pages to direct patrons to relevant NLM, NIH, and other Web resources. The more labor-intensive internal CD-ROM network was replaced with Web access to indexing and abstracting tools for onsite users and NLM staff.

In addition to those who come to use the Library's collection and onsite reference services, NLM attracts many U.S. and international visitors interested in biomedical communication, libraries, or the specific NLM exhibitions. In FY 1998, LO staff members assisted in conducting 151 regular daily tours for a total of 596 visitors. The Office of Communications and Public Liaison (formerly Office of Public Information), in the Office of the Director, made arrangements for 147 special group tours and briefings for 2,272 visitors.

National Network of Libraries of Medicine

The NN/LM works to provide U.S. health professionals, researchers, educators, administrators, and other interested people with timely, convenient access to biomedical and health information resources. The NN/LM strives to ensure equal access to up-to-date information irrespective of the user's location or institutional affiliation. The network includes more than 4,500 health sciences libraries of all sizes and types located throughout the country. LO's NN/LM Office oversees the network programs which are coordinated and administered by eight Regional Medical Libraries under contract to NLM. (See Appendix 1 for a list of the RMLs.)

Internet connectivity facilitates access to many high-quality information sources so the RMLs focus on ways to facilitate connectivity for member libraries and health professionals. As a follow-up to a 1997 survey of NN/LM member access to the Internet, NLM and the RMLs developed an "Internet Access Packet" (containing an Internet Access checklist, connection success stories, security precautions and Internet usage policies, a bibliography, and a glossary of Internet terms) intended for use by hospital librarians in discussing the need for Internet access with their administrators. This packet, along with a letter from the NLM Director and some additional supporting materials, was sent in early September to network librarians who had reported no Internet access in the 1997 survey. In late September, the Chief Administrative Officers of these institutions were sent a separate letter from the NLM Director and the brochure, *Improving Internet Access in Your Hospital: Strategic, Financial, and Clinical Benefits of the Internet*, also designed by NLM and RML staff. NLM and the RMLs plan a 1999 follow-up survey of NN/LM members who reported no Internet access in 1997 to determine their current state of connectivity.

The NN/LM is a core component of NLM's outreach program. The RMLs and other NN/LM members develop and conduct many projects to reach underserved health professionals and patients in both rural and inner city areas. Examples of the 18 special projects funded in FY 1998: the

Delaware Academy of Medicine will collaborate with the Delaware HIV Consortium to provide a path to clinical and consumer-oriented information on HIV/AIDS to all persons with HIV/AIDS, their care givers, and health care providers in the state of Delaware; George Washington University will provide Internet training to physician assistants in Maryland; the University of Puerto Rico Medical Sciences Campus Library has strengthened its computer and telecommunications infrastructure and provides technical, database, and informatics training to health professionals in Haiti, the Dominican Republic, and the Carolina Hospital in Puerto Rico; the Ohio State University Health Sciences Library and Region 5-A of the Ohio AHECs are collaborating to provide access to electronic information resources for the staffs of 6 neighborhood health centers in Columbus, Ohio; and the Idaho Health Sciences Library of Idaho State University will establish Internet access for the five Health West, Inc. Clinics, community and migrant health centers in southeastern Idaho.

The NN/LM Office works with NLM's National Information Center for Health Services Research and Health Care Technology (NICHSR) and NLM's Specialized Information Services Division to coordinate NLM's participation in *Partners in Information Access for Public Health Professionals*, a joint initiative of NLM, the NN/LM, the Centers for Disease Control and Prevention, the Health Resources and Services Division, the Association of State and Territorial Health Officials, and the National Association of City and County Health Officials. The purpose of this program is to improve access to advanced information technology and information services for practicing public health professionals. As part of this initiative, in March 1998 NLM joined the New York Academy of Medicine's Division of Public Health in co-sponsoring a national forum on *Accessing Useful Information: Challenges in Health Policy and Public Health* to obtain information on public health information needs and information access problems. LO staff members were heavily involved in planning, conducting background studies, and presenting papers at this forum. In September 1998, NLM and the NN/LM awarded 13 special outreach projects with a total budget of $650,000 aimed at various segments of the public health workforce in twelve states. NLM, the CDC, and the NN/LM are collaborating to produce a satellite broadcast directed toward public health workers on *Accessing HIV/AIDS Information Resources*, to be broadcast on February 11, 1999. NLM also has arranged for Neil Rambo, MLS, Associate Director, NN/LM Pacific Northwest Region, University of Washington, to devote 80 percent of his time in 1998/99 (under an Intergovernmental Personnel Act agreement) to study the information needs and information seeking behaviors of public health personnel in Washington state and to develop and test prototype Web information resources designed to meet some of these needs.

The introduction of free Web access to MEDLINE in 1997 dramatically increased the use of NLM online services by members of the public. During FY 1998, the NN/LM Office was heavily involved in the NLM effort to develop a pilot project to assist public libraries in providing health information services to their users. The project is a joint effort of NLM, the Friends of the NLM, the Kellogg Foundation, the Medical Library Association, and the Public Library Association, a division of the American Library Association. It involves 39 libraries or library systems (more than 200 separate facilities) in 9 states and the District of Columbia, each paired with an RML or Resource Library in the NN/LM which will provide training and other assistance. Representatives from participating libraries met at NLM in July 1998 to discuss the goals of and logistics for the project; to inform NLM about current health information services offered by participating libraries; to see a demonstration of the prototype of MEDLINEplus, NLM's new consumer health Web pages which will be tested during the pilot; and to receive initial training in the use of PubMed and Internet Grateful Med. Staff members in participating libraries will receive additional training from RMLs or resource libraries before the project is launched in October 1998. NLM will evaluate the effectiveness of the pilot and its scalability.

The RMLs and other NN/LM members conduct the majority of exhibits and demonstrations of NLM's products and services at health professional meetings around the country. LO staffs many of the exhibits in the Washington D.C. area, at the annual meetings of the Medical Library Association and some other organizations, and at some distant meetings focused on health services research, public health, and related fields. In FY 1998, NLM and NN/LM services were displayed at 81 exhibits at national, regional, and state association meetings across the U.S.

Historical Exhibitions and Programs

The History of Medicine Division mounts one major historical exhibition in the NLM lobby and rotunda each year in collaboration with the Office of the Director, the Office of Administration, and the Office of Communication and Public Liaison. FY 1998's major historical exhibition, *Frankenstein: Penetrating the Secrets of Nature*, opened on October 30, 1997. Dr. and Mrs. Lindberg hosted an opening costume party reception sponsored by the Friends of

the National Library of Medicine for the NLM staff and invited guests, including Boris Karloff's daughter. An associated film series of five classic Frankenstein movies introduced by guest speakers was presented in the Lister Hill Center Auditorium in November-December 1997. There were many special tours of the exhibit throughout FY 1998. An online version of the exhibition is available on NLM's Web site.

HMD also installs two or more "mini-exhibits" each year in the exhibit cases at the entrance to the HMD Reading Room. In FY 1998, these small exhibits were *We Were Here First: A History of the NLM Site* and the History of the Public Health Service, the latter prepared by the Public Health Service historian.

HMD, the NLM EEO Office, and the Washington Society for the History of Medicine co-sponsored a lecture by Dr. Gerard Fergerson on "The Forgotten Radicals: Health Activists and the Civil Rights Movement of the 1960s" at NLM on February 5 in observance of African-American History Month. Dr. Marilyn Yalom presented the Women's History Month lecture, "The Breast: A Cultural History" on March 23. HMD also arranged a series of 11 historical lectures for the NLM and NIH communities.

As part of a project to document the history of health services research, NICHSR arranged for oral history interviews of 13 important individuals and worked with the Lister Hill Center to videotape 9 others. Persons interviewed in FY 1998 included Morris Collen, Karen Davis, Avedis Donabedian, Charles D. Flagle, Eli Ginzberg, Edward J. Hinman, Barabara Starfield, Joseph Newhouse, Dorothy P. Rice, Barbara Starfield, John E. Wennberg, and Kerr White. NICHSR and LHC also arranged to videotape the Alice Hersh Lecture on the history of efforts to establish national health insurance at the annual meeting of the Association for Health Services Research in June 1998.

HMD staff members presented historical papers and lectures during the year and also published the results of their scholarship in books, chapters, articles, and reviews.

Training Programs for Health Sciences Librarians

In addition to the online services training programs for librarians and other search intermediaries, LO also directs the NLM Associate Fellowship program for post-masters librarians and a continuing education program in health services research and related topics for practicing health sciences librarians.

In FY 1998, LO began implementation of the expanded Associate Fellowship program announced in 1997 by recruiting 8 (rather than 4) participants for the 1998/1999 program that began in September 1998. This is the first group of Associates who will be able to compete for an optional second Fellowship year to be spent in an academic medical center, hospital, or other health-related institution. The purpose of the new second year is to give librarians experience in working with multi-disciplinary teams to integrate library and information services into the patient care, education, or research programs of a parent institution. The four 1997/98 Associate Fellows completed their one-year program in August 1998 and went on to jobs at the Arizona Health Sciences Library, the Mercer University School of Medicine Library, the New York Academy of Medicine RML Office, and the University of Wisconsin Health Sciences Libraries.

NICHSR continues to locate, develop, field-test, and teach continuing education courses for health sciences librarians in the fields of health services research and public health and to make related training materials available on NLM's Web site. At the Medical Library Association meeting in May 1998, NICHSR staff taught a class on *Introduction to Health Services Research* and arranged for a second class on *Epidemiology and Research Design* taught by Jonathan Tobin. Both courses received very favorable evaluations. *Literature Search Methods for Developing Clinical Practice Guidelines* and *TA101: Introduction to Health Care Technology Assessment* by Clifford S. Goodman, Ph.D., The Lewin Group were made available on NLM's Web site.

In celebration of the first National Medical Librarians Month in October 1997, LO and the Lister Hill Center created special Web pages highlighting the achievements of a number of health sciences librarians in National Network of Libraries of Medicine and posters featuring NLM staff librarians for display onsite at NLM’s facilities in Bethesda. Both used a baseball card motif. The onsite posters emphasized that librarians are key players on the multidisciplinary team that builds and delivers NLM services.

May 1998 – May 1999 is the centennial year of the Medical Library Association. On May 12, 1998, NLM and the MLA sponsored “The Digital Library: An Oxymoron? A Colloquium to Honor Medical Librarians” at the Lister Hill Center. The conference, which was also broadcast nationally via satellite, featured the NLM/MLA Joseph Leiter Lecture Jean-Claude Guedon, Ph.D., professor, Department of Comparative Literature, University of Montreal, and a panel of medical librarians. The panel included Judith Messerle, director, Countway Library of Medicine, Harvard Medical School, Boston, MA; Michele S. Klein, AHIP, director of library services, Medical Library, Children’s Hospital of Michigan, Detroit, and member of NLM’s Board

of Regents; Robert M. Braude, Ph.D., assistant dean, information resources, Medical Library, Cornell University Medical College, New York; and Lucretia W. McClure, librarian emerita, Edward G. Miner Library, University of Rochester School of Medicine and Dentistry, Rochester, NY. NLM also produced and distributed a poster at the MLA Annual Meeting in honor of the Centennial and made arrangements for long-term storage of MLA's Centennial time capsule.

Health Services Activities

LO continued to represent NLM and NIH on the Department of Health and Human Services (HHS) Health Data Standards Committee, which is overseeing the implementation of the administrative simplification provisions of the Health Insurance Portability and Accountability Act of 1996. In 1998, HHS published several proposed regulations for electronic health data standards required by the Act. LO contributed substantially to the codes and classifications language in the proposed rule on administrative transactions. The proposed rule on security of electronic health data was based on the recommendations in the NLM-commissioned National Research Council report, *For the Record: Protecting Electronic Health Data* (1997).

LO staff members serve as project officers on telemedicine evaluation contracts with Brigham and Women's Hospital, Columbia University, Georgetown University, Indiana University, the University of Pittsburgh, and the University of Washington. These projects are funded by NLM's Office of High Performance Computing and Communications in the Lister Hill Center.

Table 1

Growth of Collections

Collection	*Previous Total (9/30/97)*	*Added FY 1998*	*New Total (9/30/98)*
Book Materials			
Monographs:			
Before 1500	578	0	578
1501-1600	5,804	7	5811
1601-1700	10,127	2	10,129
1701-1800	24,453	19	24,472
1801-1870	41,131	16	41,147
Americana	2,341	0	2,341
1870-Present	644,513	14,924	659,437
Theses (historical)	281,794	0	281,794
Pamphlets	172,021	0	172,021
Bound serial volumes	1,101,264	28,437	1,129,701
Volumes withdrawn	(60,810)	(5,671)	(66,481)
Total volumes	2,223,216	37,734	2,260,950
Nonbook Materials			
Microforms:			
Reels of microfilm	89,998	5,922	95,920
Number of microfiche	372,309	24,271	396,580
Total microforms	462,307	30,193	492,500
Audiovisuals .	62,083	1,699	3,782
Computer software	1,167	241	1,408
Pictures	56,601	0	56,601
Manuscripts	2,454,542	274,530	2,729,072
Total nonbook	3,036,700	306,483	3,835,863
Total book and nonbook	**5,259,916**	**344,217**	**5,797,725**

Table 2

Acquisition Statistics

Acquisitions	*FY 1996*	*FY 1997*	*FY 1998*
Serial titles received	22,522	22,378	22,247
Publications processed:			
Serial pieces	154,146	150,713	146,921
Other	20,846	21,422	21,642
Total	174,992	172,135	168,563
Obligations for:			
Publications	$5,012,058	$5,282,348	$5,266,996
Included for rare books	($209,178)	($267,221)	($251,293)

Table 3

Cataloging Statistics

	FY 1996	*FY 1997*	*FY 1998*
Completed Cataloging	20,365	20,296	18,803

Table 4

Bibliographic Services

Services	*FY 1996*	*FY 1997*	*FY 1998*
Citations published in MEDLINE*	322,825	519,012	411,921
For *Index Medicus**	309,038	499,794	388,022
Recurring bibliographies	6	----	----
Journals indexed for *Index Medicus*	3,205	3,211	3,302
Abstracts entered	242,544	398,576	312,064

** Figures for FY 1997 reflect reduction of a backlog carried over from FY 1996.*

Table 5

Circulation Statistics

Activity	*FY 1996*	*FY 1997*	*FY 1998*
Requests Received	584,738	630,190	694,281
Interlibrary Loan	347,992	353,408	374,791
Onsite	228,113	276,782	319,490
Requests Filled:	410,732	463,011	523,081
Interlibrary Loan	227,810	252,830	275,588
Photocopy	215,461	239,575	264,301
Original	10,814	11,670	10,167
Audiovisual	1,535	1,585	1,120
Onsite	182,922	210,181	247,493

Table 6

Online Searches—All Databases*

	FY 1996	*FY 1997*	*FY 1998*
PubMed			88,000,000
Internet Grateful Med			12,000,000
ELHILL			4,000,000
Total	7,392,947	**21,111,426	104,000,000

**Beginning in FY 1998, all figures are rounded.*
***Includes both PubMed (June 26- September 30, 1997) and ELHILL.*

Table 7

Reference Services

Activity	*FY 1996*	*FY 1997*	*FY 1998*
Reference Section:			
Offsite requests	24,799	21,297	27,070
Onsite requests	34,796	40,851	43,782
Total	59,595	62,148	70,852

Table 8

History of Medicine Activities

Activity	*FY 1996*	*FY 1997**	*FY 1998*
Acquisitions:			
Books	188	76	108
Modern manuscripts	163,500	750,000	274,530
Prints and photographs	183	306	849
Historical audiovisuals	108	36	94
Processing:			
Books cataloged	351	558	193
Modern manuscripts cataloged	0	0	0
Pictures cataloged	0	0	0
Citations indexed	1,363	2,430	1,516
Public Services:			
Reference questions answered	12,885	13,701	12,387
Onsite requests filled	5,289	6,313	3,733

**Some figures have been corrected.*

Specialized Information Services

Melvin Spann, Ph.D.
Associate Director

The Toxicology Information Program, now known as the Toxicology and Environmental Health Information Program (TEHIP), was established at NLM more than 30 years ago as the Division of Specialized Information Services (SIS). Over the years TEHIP's evolution has kept pace with the increasing need for toxicological and environmental health information by taking advantage of new computer and communication technologies. Such mechanisms have enabled us to provide more rapid access to a wider audience. Our development of novel search capabilities means that users need not have extensive search knowledge and thus allows data to be relayed to them more efficiently. Finally, we are moving beyond the bounds of the physical Library, 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. Currently, refinements and additions are being made to allow easy access to a range of information collected by this Division. Most significantly, we took the first steps in making all our toxicology data free over the Internet. Access to NLM's AIDS/HIV information resources is already free.

In FY 1998 SIS reexamined the scope and coverage of its current programs, proposed new opportunities to enhance information services, and investigated emerging areas, including the application of new multimedia technology in the delivery of information services. This examination was initiated using the mechanism of an Institute of Medicine (IOM) evaluation of the TEHIP Program, and a report from this study, *Toxicology and Environmental Health Information resources: The Role of the National Library of Medicine*, was released in the spring of 1997. Another IOM committee was convened in FY 1998 to prepare a follow-on report on Internet access to the toxicology-related resources of NLM.

Resource Building

The wide range of resources related to toxicology and environmental health information and AIDS/HIV 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 6,000 searches each month. Increased emphasis was placed on providing more data on human toxicology and clinical medicine within HSDB, in keeping with earlier recommendations of the Board of Regents Subcommittee on TEHIP. Changes to the composition of the Scientific Review Panel (SRP) were made to accommodate the shift in content emphasis. Newer sources of relevant data are being examined for incorporation into new and existing data fields within the current 4,533 HSDB records. Because of increased staff efforts, more records are being processed through special enhancements, including source updates from various peer-reviewed files. These enhancements are being made possible by a customized Windows-based PC workstation with enhanced file-building features.

CHEMID (Chemical Identification File) is an NLM online chemical dictionary containing over 344,000 records, primarily describing chemicals of biomedical and regulatory importance. It also contains an important set of regulatory data, collectively known as Supersets. Over 13,000 records are augmented with the name and an indication of source for chemicals mentioned in one or more of 31 lists, e.g., the Department of Transportation Hazardous Materials List, and the Priority List of the Agency for Toxic Substances and Disease Registry (ATSDR). These data allow users to determine if a chemical is mentioned on a given list and under what name, as well as to search for chemical classes on these lists. During FY 1998, an extensive quality control effort was made on data in this file and new nomenclature data was added to enrich the file content.

CHEMLINE (Chemical Dictionary Online) an online chemical dictionary and directory file that was primarily supplied by the Chemical Abstracts Service (CAS) and required extensive royalty payments was dropped as an NLM database in January 1998 as part of the move to provide free access to all NLM databases.

TOXLINE (Toxicology Information Online) is an NLM online bibliographic retrieval service produced by merging "toxicology" subsets from some eighteen secondary sources. TOXLINE and its backfile, **TOXLINE65**, contain data from sources that do not require royalty charges based on usage. Additional information from the Chemical Abstracts Service requiring usage royalties was contained in two separate online bibliographic files, **TOXLIT** and **TOXLIT65**, which were discontinued in January 1998 as part of the move towards free access to NLM databases. At the end of FY 1998 TOXLINE and its backfile contained over two and one half millions citations.

This decision to no longer carry the TOXLIT files and the move to free access has prompted us to explore alternative ways to acquire and build and organize toxicology bibliographic information.

DIRLINE (Directory of Information Resources Online) is NLM's online directory of resources including organizations, databases, bulletin boards, as well as projects and programs with special biomedical subject focus. These resources provide information to users that may not be available from one of the other NLM bibliographic or factual databases.

The availability of DIRLINE via the Internet through NLM's Locator and on the web continues to result in a high level of use of the database and a higher level of recognition of its utility by biomedical librarians, health professionals, and the public. Increased funding for database maintenance has made it possible to improve the quality and timeliness of the content of the database and improve collaboration with other subfile producers. An online version of *Health Hotlines*, NLM's popular publication listing toll-free telephone numbers for health-related sources, is now available on the Web.

The **Toxic Chemical Release Inventory (TRI)** series of files now includes 10 online files, TRI87 through TRI96. These files remain an important resource for environmental release data and continue to attract new users. Mandated by the Emergency Planning and Community Right-to-Know Act (Title III of the Superfund Amendments and Reauthorization Act of 1986), these EPA-sponsored databases contain data on environmental release data to air, water, and soil for over 600 EPA-specified chemicals. The TRI95 file, released in May 1997, included over 300 additional chemicals, as well as required submissions from federal facilities and military installations.

TRIFACTS, a companion file to the TRI series, supplies users with information related to health and ecological effects and the safety and handling of the initial TRI chemicals. These records are supplied by EPA, who in FY98 continued to also fund the management and maintenance of the TRI files on TOXNET.

The **Chemical Carcinogenesis Research Information System** (CCRIS) continues to be built, maintained, and made publicly accessible by NLM. This data bank is supported by the National Cancer Institute (NCI) and has grown to 7,786 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 with the user community. 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 receives a high level of interest among users in other countries.

The **Registry of Toxic Effects of Chemical Substances** (RTECS) is a data bank based upon a National Institute of Occupational Safety and Health (NIOSH) file by the same name which NLM has restructured and made available for online searching. SIS continues to add new data to this file as NIOSH makes them available. Now that NIOSH no longer prints the full RTECS, online access via NLM to this widely used resource including over 142,000 chemical records has become even more important.

The **Developmental and Reproductive Toxicology** (DART) database now contains over 34,000 citations from literature published since 1989 on agents that may cause birth defects. Records from DART are also added to TOXLINE on a quarterly basis. DART is a continuation of the Environmental Teratology Information Center backfile (ETICBACK) database, which contains almost 50,000 citations to literature published from 1950-1989. ETICBACK citations are also found in TOXLINE. DART is funded by NLM, the Environmental Protection Agency, the National Institute of Environmental Health Sciences and the FDA's National Center for Toxicological Research and is managed by NLM.

The **Environmental Mutagen Information Center** (EMIC) database contains over 20,000 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. Records from EMICBACK are included in TOXLINE. Plans are under way to add the records from the new EMIC database to TOXLINE as well. EMIC is funded by the Environmental Protection Agency and the National Institute of Environmental Health Sciences and managed by NLM.

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 (http://sis.nlm.nih.gov) users can access interactive retrieval services in toxicology and environmental health or AIDS/HIV information, find program descriptions and documentation, or be connected to outside related resources.

Toxicology Data Network (TOXNET)

The **Toxicology Data Network (TOXNET)**, NLM's computer system providing access to the majority 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. Concurrent with this transition to improved hardware and software was the announcement that TOXNET access would be free of charge via an easy-to-use World Wide Web interface (http://toxnet,nlm.nih.gov). The files have been logically organized for searching into three groups: toxicology data, toxicology literature, and toxic releases. Many options are available for searching and displaying results, which can also be e-mailed or FTP'd to users. In addition, TOXNET's web interface includes links to PubMed and Internet Grateful Med, the latter now offering free access to TOXLINE and ChemID. Work is proceeding on further refining and enhancing the interface through a more intuitive search engine, ranked relevancy listings, improved navigational features, and cross-file capabilities. Consideration is also being given to consolidating access to TOXLINE and ChemID into the TOXNET system.

Internet Grateful Med (IGM)

Near the end of FY 1998 access to TOXLINE and ChemID was added to **IGM,** where access to DIRLINE, the AIDS/HIV databases, MEDLINE, and many other NLM databases was already being provided. This route of access (http://igm.nlm.nih.gov) is free to all users.

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 a includes a prototype database for construction of ChemID records. It is expected to evolve into the primary chemical dictionary for users of the SIS databases.

AIDS Information Services

The recommendations from the 1993 NIH HIV/AIDS Information Services Conference remain the guide for NLM's HIV/AIDS information programs and services. In 1996 NLM initiated a World Wide Web homepage for AIDS (http://sis.nlm.nih.gov/aids/). This distribution mechanism has enhanced NLM's ability to get important information to the people who need it. In addition to links with NLM's online databases, fact sheets and publications, links to other NIH components are available as part of an new online version of the *Guide to NIH HIV/AIDS Information Services*. New databases of abstracts presented at important AIDS conferences have also been made available through this mechanism prior to their addition to AIDSLINE.

Two collaborative projects with other PHS agencies, the AIDS Clinical Trials Information Service (ACTIS) and the HIV/AIDS Treatment Information Service (ATIS) continue to be very successful. Many other information providers are downloading data from AIDSTRIALS and AIDSDRUGS and adding it to their services. A number of World Wide Web sites are providing the data to their users in a number of formats and the data have also been distributed internationally through this mechanism. In addition, in 1997 AIDSTRIALS and AIDSDRUGS became available for searching via Internet Grateful Med and via a Web interface at ACTIS.

NLM initiated a fifth round of AIDS Outreach Purchase Orders in FY 1998 and made awards to 19 community-based organizations and libraries. Seven previously funded projects that had shown evidence of success received funding to continue or expand their activities.

NLM has continued to work with the Health Information Center at the Wheaton Regional Library (Md.) to complete the collaborative project in AIDS and toxicology information. Outreach was expanded through satellite centers at three other regional libraries. Wheaton Library staff developed and conducted training for the public in the use of the World Wide Web for information retrieval. Training and outreach to the Historically Black Colleges and Universities has resulted in the development of training materials with culturally specific content. A number of training sessions were held during the year with very positive feedback from the trainees.

Outreach/User Support

SIS continues its support of the Toxicology Information Outreach Project. The objective of this initiative is to strengthen the capacity of Historically Black Colleges and Universities (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 nine HBCUs participating in a pilot training development project, NLM has collaborated with the Agency for Toxic Substances and Disease Registry (ATSDR) to train representatives from 61 additional schools in the use

of NLM's valuable online resources. One of the training classes, hosted by Texas Southern University, included HBCUs and community-based organizations from the Lower Mississippi Delta. This class was jointly sponsored by ATSDR, NLM, and the Environmental Justice Office of the Environmental Protection Agency in support of the Mississippi Delta Project. Other classes with specific user group focus have been conducted in addition to our usual NLM-based training. In addition, a bibliography on environmental justice was compiled and is available on SIS's WWW server, along with other bibliographies and written user resources.

User Support Computer-Based Activities

In FY 1998, SIS developed of TOXICOLOGY TUTOR II, second in a series of introductory level computer-based toxicology courses. It takes advantage of Windows' graphical user interface incorporating color graphics, photography, animation, and audio. The TOXICOLOGY TUTOR series is designed for the Internet, and is available on the SIS web server.

Additionally, SIS has developed a computer-based demo for the Internet illustrating the TEHIP databases, ChemID, TOXLINE, RTECS, HSDB, CCRIS, IRIS, TRI, TRIFACTS, GENETOX, EMIC, and DART. It was updated in FY 1998, along with an Internet version of the TEHIP slide overview.

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 Internet through the SIS WWW server.

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 Congress in 1968 and serves as an intramural research and development division of the National Library of Medicine. The Center conducts research, provides training opportunities, develops and deploys systems used in NLM production activities, and sponsors external research through research contracts in telemedicine and medical applications for the Next Generation Internet.

Center research is carried out through several major programs, all sharing the purpose of improving health-care information dissemination and use. Center staff are engaged in developing intelligent information storage and retrieval systems, conducting their research by drawing on a diverse set of scientific fields and methods. Current staff have backgrounds in medicine, computer science, library and information science, linguistics, cognitive science, education, and engineering. Research projects include medical knowledge representation, digital library research, automated indexing techniques, vocabulary and thesaurus research, natural language processing, image processing, database design, machine learning, expert systems, and computer-based learning. The Unified Medical Language System knowledge sources, the Visible Human data set, and MEDLINE are important sources of data for many experiments. The most current information about Center activities can be found at the Center's Web site (http://lhncbc.nlm.nih.gov/). The Center is administratively organized into several components, although many research projects involve collaboration across these organizational units. A brief synopsis of the major activities of each of the components follows.

As its most significant effort, the *Audiovisual Program Development Branch* supports the Center's research, development, and demonstration projects with high quality video, audio, imaging, and graphics materials. From initial project concept, through final project implementation, evaluation and reporting, including image creation, preservation, transfer and display, all forms and formats of imaging are supported. Consultation and materials development are also provided by the branch for the NLM's educational and information programs. With the mission requirement of the Library expanded to include effective outreach activities, the support that the branch provides to these programs continues to increase. From the use of optical media technologies to teleconferencing and World Wide Web design support, the graphics, video, and audio materials requirement has increased in quantity and diversified in format. Another area of concentration is the engineering of technical improvements applied to issues such as image quality and resolution, color fidelity, media transportability, media storage, and image communication. High resolution video is an 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 cost advantages that they offer. Three dimensional computer graphics, animation techniques, and photorealistic rendering methods have changed the tools and products of the graphics artists in the branch. Digital video and image compression techniques are central to projects being pursued to improve image storage and transmission. Information about Audiovisual Program Development Branch projects appears on the Branch Web server at http://lhncbc.nlm.nih.gov/apdb/.

The *Cognitive Science Branch* conducts research and development in computer and information technologies, disseminates information about these technologies to the NLM's various constituencies, and supports their application in health professions education. The Learning Center for Interactive Technology, a part of the branch, is a place where individuals come to learn about the most current technology and its potential application in the health care setting. Medical language processing is an important research area for the branch and involves linguistic and knowledge-based methods for improving access to complex biomedical information. Branch members collaborate with other research staff in a project to develop automated and semi-automated techniques for indexing the biomedical literature. Branch staff participate in most aspects of the Unified Medical Language System project, including development and maintenance of the UMLS Knowledge Source Server, the Semantic Network, and the SPECIALIST lexicon. The branch conducts research in digital libraries and is currently collaborating with NLM's History of Medicine Division on *Profiles in Science*, a project to digitize collections of prominent biomedical scientists. Information about Cognitive Science Branch projects appears on the Branch Web server at http://wwwcgsb.nlm.nih.gov/.

The *Communications Engineering Branch* conducts research in image engineering and communications engineering, motivated by NLM's

mission-critical tasks such as document delivery, archiving, and automated data entry, as well as wide Internet access to mixed text and x-ray image databases, and future imaging applications in support of medical educational packages employing digitized radiographic, anatomic, and other imagery. The Branch conducts R&D in the capture, storage, processing, online retrieval, transmission and display of both biomedical documents and medical imagery. Data types of interest include bitmapped bitonal document images, digitized color documents, digitized x-rays, color cryosection images from the Visible Human collection, and motion video. Areas of active investigation center on image compression, image enhancement, image understanding, pseudo-grayscale rendition, image transmission and networks implemented via asynchronous transfer mode (ATM) and satellite technologies, optical character recognition (OCR) and man-machine interface design applied to automated data entry. Information about Communications Engineering Branch projects appears on the Branch Web server 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. Branch projects involve both basic and applied research in such areas as expert systems, intelligent database systems, multimedia hypertext information delivery, machine learning, data mining, and machine-assisted indexing for information classification and retrieval. Research issues include knowledge representation, knowledge base structure, knowledge acquisition, and the human-machine interface for complex systems. Important components of the research include embedded intelligence systems that combine local reasoning with access to large-scale online data banks. Other project work involves multimedia knowledge-based systems with interactive video capability and systems that search multiple databases with a single user query. Branch staff participate in the Unified Medical Language Systems project, including the development of the largest of its knowledge sources, the UMLS Metathesaurus. Branch staff coordinate medical informatics training, including the eight-week NIH elective in medical informatics for third-year and fourth-year medical students held each spring. Information about Computer Science Branch projects appears on the Branch Web server at http://lhncbc.nlm.nih.gov/csb/.

The *Office of High Performance Computing and Communications* serves as the focal point for NLM's High Performance Computing and Communications planning, research and development activities with federal, industrial, academic, and commercial organizations. The major activities of the Office involve the Visible Human Project, telemedicine contracts, and the Next Generation Internet. 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. Next Generation Internet capabilities will be required if this emerging infrastructure is to 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, and infrastructure technology as a means for collaboration. Information about the Office of High Performance Computing and Communications projects appears on their Web server at http://lhncbc.nlm.nih.gov/ohpcc/.

Research Programs

Center programs and activities involve a number of research areas:

- Computer and Information Science Research, including medical language processing, expert systems, machine learning, and Web-based application development
- Image Processing Research, including image acquisition, storage, retrieval and dissemination, and research in digital libraries
- Education and Training, including medical informatics training, learning technologies, and video, audio, and graphics support for effective outreach activities
- Resource Support and Development, including core system support, system security and advanced network planning
- External Research Support, including support for telemedicine and medical applications for the Next Generation Internet

Computer and Information Science Research

Unified Medical Language System (UMLS)

UMLS Metathesaurus

The Metathesaurus is the largest of the UMLS knowledge sources. It is a machine-readable knowledge source representing multiple biomedical vocabularies organized as concepts in a common format. It thus provides a rich terminology resource in which terms and vocabularies are linked by meaning. In 1998, the UMLS Metathesaurus group

continued its two main tasks: producing increasingly larger annual editions of the Metathesaurus with new and updated vocabulary sources, and developing and deploying new software systems for work on concept-oriented terminologies.

The 1999 version of the Metathesaurus will again be augmented by more than 100,000 additional concepts. This increase of more than 25 percent was accomplished while moving to a completely redesigned software environment for Metathesaurus production. Major improvements of the new software include increased efficiency, full records of all editing actions, and "undo" capabilities. The software team has also added integrated editing management and workflow management and is developing a "subsetter" that provides UMLS users with a straightforward means of excluding from their local copy of the Metathesaurus any constituent vocabularies for which they do not have appropriate license agreements or which they otherwise do not want to include in their local application. Working closely with the MeSH Section and the PubMed and Internet Grateful Med development groups, the Metathesaurus team has developed and deployed another system which provides Metathesaurus synonymy for PubMed and Internet Grateful Med searching. This system includes mechanisms which allow the MeSH Section immediately to add or modify synonymy for retrieval while working on changes to MeSH.

In the next fiscal year the UMLS Metathesaurus group will further extend and test software for automated quality assurance and automated production of future Metathesaurus releases. The development of remote client-server operations will continue, in parallel with the development and evaluation of alternative Java-based platform-independent editing and management systems. Finally, the group will continue to emphasize means of developing and maintaining the Metathesaurus involving multiple autonomous terminologies and a variety of collaborating institutions.

UMLS Information Sources Map

The Information Sources Map (ISM) project demonstrated a third-generation prototype for Sourcerer, a system capable of accepting a natural language query, identifying information sources capable of answering the query, and retrieving relevant data from those sources. The system included the ELVIS Java ISM client and Java front-ends to virtually all of NLM's information sources as well as the Natural Language Systems group's MetaMap application and the ISM database server. These components employed CORBA (Common Object Request Broker Architecture) as the means of sharing data objects over a network. In light of the rapid evolution of Web-based information retrieval, the work of the ISM team will be carried forward into a new project called "Emerging Network Retrieval Protocols."

UMLS SPECIALIST Lexicon and Lexical Programs

The SPECIALIST lexicon is a large syntactic lexicon of medical terminology. Lexical items are collected into unit records containing morphological, syntactic and spelling information about each item. The lexicon contains over 125,000 lexical items in the 1999 release of the UMLS. Morphological information includes full inflectional information about each item. A data base of derivational relationships is also maintained with the lexicon. Morphological and spelling information is important for matching of lexical items and forms a major part of the capability of the lexical tools. Syntactic information includes verb complement patterns for verbs and sequencing information for adjectives. Lexicon building tools have been used to facilitate lexicon building and maintain the consistency of the lexicon. These tools enforce a complete and consistent form for lexical entries, and provide users with a menu based approach to entering lexical information. A lexicon grammar specifies the correct form of lexical records and assures that lexical records created by the tools are correctly formed. The lexicon building emphasizes quality control and correction of errors in the existing lexicon as well as growth. Current efforts have included a review of spelling variation within the lexicon and a review of the morphological databases. A word list derived in part from the SPECIALIST lexicon is now used in the MARS system for semi-automatic entry of journal article data into MEDLINE.

The lexical programs are designed to help users abstract away from certain sorts of lexical variation. Retrieval and matching tasks often require some way to ignore variation. The programs use the SPECIALIST lexicon and associated databases along with a set of heuristic rules to allow users to abstract away from alphabetic case, word order, variations in punctuation, spelling variation, inflection and derivation. The programs offer methods (called flows) which can combine these functions to produce indexes with varying degrees of aggressiveness with respect to the matching of terms. The programs are used by the Metathesaurus to create the normalized string index and to create the normalized word index. During this past year a JAVA version of the lexical programs was written and this may eventually replace the C version.

UMLS Semantic Network

The purpose of the Semantic Network is to provide a consistent categorization of all concepts represented in the Metathesaurus and to provide a set of useful relationships between these concepts. All information about specific concepts is found in the Metathesaurus; the Network provides information about the set of basic semantic types, or categories, which may be assigned to these concepts, and it defines the set of relationships that may hold between the semantic types. The 1999 release of the Semantic Network contains 134 semantic types and 54 relationships. The Semantic Network serves as an authority for the semantic types that are assigned to concepts in the Metathesaurus and defines the types, both with textual descriptions and by means of the information inherent in its hierarchies. The semantic types are the nodes in the Network, and the relationships between them are the links. There are major groupings of semantic types for organisms, anatomical structures, biologic function, chemicals, events, physical objects, and concepts or ideas. The current scope of the UMLS semantic types is quite broad, allowing for the semantic categorization of a wide range of terminology in multiple domains.

UMLS Knowledge Source Server

The UMLS Knowledge Source Server is an evolving tool for providing Internet access to the information stored in the UMLS Knowledge Sources. The purpose of the Knowledge Source Server is to make the UMLS data more accessible to users, and in particular to system developers. The system architecture is based on the client server paradigm wherein remote site users send their requests to a centrally managed server at the NLM. The client programs can run on any platform supporting the TCP/IP communication protocol. Access to the system is provided through a command-line interface, through an Application Programming Interface (API), and through the World Wide Web. The Web interface allows users to browse and explore the data and to see how those data are organized in the UMLS. Information about Metathesaurus concepts, semantic types and relations in the Semantic Network, and lexical items in the SPECIALIST lexicon can be found by querying the system. The command-line interface is best suited for batch processing. Researchers can submit a list of terms to the server and retrieve a range of information about those terms, and they can further filter the results, limiting the result set, for example, by a particular attribute. The API allows developers at remote sites to embed calls in their application programs to the Knowledge Source Server, thereby accessing the UMLS data directly over the Internet.

Natural Language Processing

Modular Text Processing

The Lexical Systems group is developing a text analysis system based on the SPECIALIST lexicon and the lexical tools. This system is modular in design to allow for flexible use and continuous revision. The modules are servers which will be available to a variety of clients for a variety of uses. Interchangeable versions of particular modules should facilitate experiments to improve the system. The system consists of several modules: a tokenizer module to analyze text into tokens and label them; a sentence identification module to analyze text into sentences; a lexical look-up module to find lexical items in the text, a shapes module to identify items in the text that do not occur in the lexicon but have types recognizable from their form, and a parser module to assign phrase structure to the sentences of the text. A structured text analysis module is being developed. This module deals with structured text such as text with SGML markup, MEDLINE record structure or other structuring devices. The tokenizer and sentence identifier modules in this version use regular expressions to break text into tokens and sentences and to identify the type of each token. The pattern used to identify sentence breaks was selected after some preliminary empirical studies of sentence breaks in MEDLINE text. The lexical look-up module uses a Berkeley DBMS database. Multi-token lexical items (multi-word terms) are retrievable by their first token so the lexical look-up module can detect multiple token lexical items without combinatorial explosion.

The design of the parser module is strongly based on the syntactic information encoded in the SPECIALIST lexicon. After consideration of several grammar paradigms, Categorical Grammar was chosen as a grammatical system that expresses this lexically based approach to sentence structure. The lexical record for each lexical item can be translated into a categorical grammar category symbol which expresses how that lexical item combines with other lexical items and phrases to form a larger phrase. A prototype version of the Categorical Grammar Parser has been developed and demonstrated. Experiments on noun phrase extraction using the Categorical Grammar Parser are being designed. The parser will be used to extract noun phrases from a body of MEDLINE text, and the results will be hand corrected to establish "ground truth" for experiments comparing the parser to other noun phrase extraction schemes.

Semantic Knowledge Representation

The Semantic Knowledge Representation project is concerned with reliable and effective management of the knowledge encoded in natural language texts. The project develops programs which provide usable semantic representation of biomedical text by building on currently available resources, especially the UMLS knowledge sources and lexical processing tools. Two programs in particular, MetaMap and SemRep, are being evaluated, enhanced, and applied to a variety of problems in the management of biomedical information. These include automatic indexing of MEDLINE citations, concept-based query expansion, analysis of complex Metathesaurus strings, accurate identification of anatomical terminology and relationships in clinical records, and the mining of biomedical text for chemical binding relations.

Recent enhancements to the technology underlying the Semantic Knowledge Representation project include the parallelization of the MetaMap program and the introduction of a hierarchical tokenization scheme for the initial stages of MetaMap processing. The parallelization effort uses Parallel Virtual Machine (PVM), a software package that provides a unified framework for developing parallel programs using heterogeneous machines on a network. An additional software package called Condor allows for using non-dedicated machines effectively without impeding the machines' primary tasks. In recent tests of PVM and Condor, MetaMap processing is six times faster than before. The second effort is the introduction of hierarchical tokens into MetaMap's initial text processing algorithm. Two types of hierarchical tokens, author-defined acronyms or abbreviations and complex chemical names, have been implemented and increase MetaMap's accuracy by restricting the context in which some text is analyzed. Future efforts will be devoted to elucidating genetic terminology, complex numeric expressions, scientific nomenclature and bibliographic citations. The effectiveness of the SemRep program has been enhanced by expanding the set of rules which map from syntactic phenomena to Semantic Network relations.

MetaMap and SemRep together have served as the basis for several ongoing projects which explore the application of focused semantic interpretation to problems in biomedical information management. One such project explores the formative evaluation of a program for accurately identifying terminology associated with the coronary arteries as expressed in coronary catheterization reports. In another, two experiments were conducted which assess the ability of an enhanced version of SemRep to identify and characterize physico-spatial semantic relationships in these same reports. A separate application addresses the ability of MetaMap and a special version of SemRep to discover molecular binding relations in biomedical text.

Indexing Initiative

The objective of the indexing initiative project is to investigate methods whereby automated indexing methods partially or completely substitute for current indexing practices. The project assumes the continued existence and growth of NLM's MeSH vocabulary and of the UMLS Knowledge Sources and is investigating concept-based indexing methods that go well beyond automatic word-based indexing. One method seeks to discover semantic relationships between phrases in text as a way of more accurately representing content. Another identifies Metathesaurus concepts in biomedical text and then maps these through a weighting algorithm based on linguistic and knowledge-based techniques to appropriate MeSH terms. One set of experiments has demonstrated the value of such concepts for automatic query expansion. Another investigation provides an enhanced representation of semantic content by ranking concepts assigned to MEDLINE abstracts on the basis of frequency of occurrence and specificity as measured by hierarchical depth in MeSH. Methods for providing better access to MeSH terminology through the UMLS are also being investigated and these should lead to improved MeSH validation procedures as part of the indexing process. A prototype indexing system is under development and is based on the results of a variety of experiments conducted during 1998.

Digital Library Research

The digital library research program investigates all aspects of creating and disseminating digital collections including proposed and adopted standards, emerging technologies and formats, effects on previously established processes, and protection of original materials. The project builds on work already accomplished under a special joint project with the History of Medicine Division begun in the spring of 1997, as well as on earlier work conducted under the EIDOS document management project. The research involves all aspects of creating and disseminating digital collections including proposed and adopted standards, emerging technologies and formats, copyright and legal issues, effects on previously established processes, protection of original materials, and permanent archival of digital surrogates. The project explores long term solutions for digital archives, new methods for creating and accessing digital library collections, the development of modular and open information environments, investigation of the role of well-structured metadata,

and the exploration of different "points of view" on the same underlying data set.

We began a project in the spring of 1997 whose goal it is to make the archival collections of prominent biomedical scientists available on the World Wide Web. The site is designed for scientists, scholars, and students, all of whom may gain an appreciation of the history of early scientific discoveries, and also share in the excitement of the scientific enterprise. The collections have been donated to the NLM and contain published and unpublished materials, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audio tapes and other audiovisual materials. In the fall of 1998 the *Profiles in Science* Web site was launched (http://profiles.nlm.nih.gov/). The first collection featured on the site is a special collection of materials relating to the work of Dr. Oswald Avery (1877 - 1955), one of this country's first molecular biologists, whose findings proved that the genetic material is DNA. A good part of 1998 was devoted to the conversion and description of the papers of Nobel Laureate Dr. Joshua Lederberg. Dr. Lederberg's collection is testing the limits of each stage in the process of digital library creation.

Expert Systems Program

The objective of the Expert Systems program is to build computer programs that facilitate access to knowledge. Such programs could serve as surrogates for the human expert or as high level tools for the expert in the presence of information overload. The Expert Systems program builds intelligent multi-media expert systems to investigate issues in knowledge representation and knowledge base structure, knowledge acquisition and knowledge base maintenance, the evaluation of knowledge-based systems, and the creation and delivery of knowledge-based systems over the Internet. Research staff have built a multimedia expert system shell called CTX, for Criteria Table Expert. The shell includes a knowledge base compiler, a run-time system which can access information from multiple knowledge sources, a knowledge base editor, a case editor and a suite of automated testing programs for analyzing system performance against sets of benchmark test cases. It has been tested successfully with knowledge bases in several widely varying domains. CTX for MS-DOS machines, WinCTX for machines running Microsoft Windows, and a CTX engine which runs on Unix and DOS machines, are complete and have been used by several outside collaborators. Internet-CTX, a version of the CTX program that delivers criteria table based applications over the Internet, is complete and has been used to deliver the AI/RHEUM diagnostic system over the Internet. Design and development of new applications using this shell will be explored in the coming year.

The AI/RHEUM knowledge-based diagnostic system in rheumatology has been a research testbed for the Expert Systems program for many years. Collaborating physician Dr. Balu Athreya, a pediatric rheumatologist, enhanced the AI/RHEUM knowledge base for more effective use with pediatric patients. Dr. Athreya's evaluation of the resulting system using 94 consecutive patients seen in a pediatric rheumatology clinic in Pennsylvania will appear in the journal Pediatrics in October 1998. He is currently completing sensitivity and specificity studies on each of the diseases represented in the pediatric version of the AI/RHEUM knowledge base. Dr. Athreya has received funding from DuPont Hospital for Children to develop a new pediatric system using the CTX programming environment.

Machine Learning Project

Machine Learning (ML) encompasses a wide variety of mechanisms for creating computer programs that automatically acquire knowledge and put it to work. The rapid growth of biomedical data at all levels, from molecular sequences to health care, challenges practitioners to find pearls of wisdom in a vast sea of data. While efficient databases make it possible to find specific cases, instances, or facts, the goal of machine learning is to use large amounts of data to produce significant general insights. Some ML algorithms, called "unsupervised learning," can cluster large amounts of data, to identify related groups that may not have been apparent on initial inspection. Other ML techniques, called "supervised learning," attempt to generate mappings from one set of data to another, often trying to find relationships between easily measurable variables and hard to predict outcomes. Yet other ML approaches, called "reinforcement learning," try to generate sequences of actions that lead to rewards of some kind.

The Machine Learning group has been active both in exploring new machine learning techniques and in applying these technologies to important problems in biomedicine. A novel machine learning system developed by this project is the ability to synergistically integrate a wide variety of supervised learning algorithms, getting performance at least equal to the best of them, and often significantly better than any of its component learners. This framework, called COEV, is the subject of a provisional patent. It is being licensed by commercial entities both inside of biomedicine and out. In addition, the Machine Learning Project has signed a letter of intent to establish a cooperative research and development agreement (CRADA) with VIPS systems of Towson, Maryland. The goal of the

CRADA is the development and application of COEV to large scale healthcare information processing systems with the intention of automatically discovering ways to control healthcare costs and reduce Medicare fraud. Recent publications have concentrated on the application of COEV and other locally developed technologies to macromolecular sequence analysis. One recent project investigated problems in the assessment of enzyme function from sequence. It described a machine learning approach to significantly improving predictive accuracy, mainly by avoiding false positive sequence matches. Another recent project investigated applications of unsupervised classification approaches to gene expression data generated by the new gene chip technology. Project staff are applying machine learning techniques to the MARS project for automated extraction of bibliographic information from paper journals. The project is also exploring the possibility of combining text-based machine learning with natural language processing to develop a database of molecular binding affinities directly from databases of scientific article abstracts.

Internet Grateful Med

Released in April 1996 as a principal component of NLM's System Reinvention initiative, Internet Grateful Med (IGM) provides users with assisted searching in multiple NLM databases over the World Wide Web. The original version searched only in MEDLINE; AIDSLINE, HealthSTAR, and the new PREMEDLINE databases were added in 1996. In 1997 seven more databases (AIDSDRUGS, AIDSTRIALS, DIRLINE, HISTLINE, HSRPROJ, OLDMEDLINE and SDILINE) were added, making a total of 11 of NLM's databases available to the IGM user. IGM is an intelligent gateway system designed to provide assisted searching to NLM's users with a consistent look and feel across multiple database systems as NLM's primary retrieval systems evolve. The gateway architecture has proven a successful means of transparently connecting users to several different types of retrieval systems while insulating them from the specifics of differing command languages. For its approach to the problem of providing attractive Internet-accessible interfaces to legacy systems, Internet Grateful Med won one of several Federal Showcase Site awards at the Federal Webmaster Workshop in 1996.

The Internet Grateful Med team developed and deployed several significant capabilities in addition to the IGM retrieval system gateway. In June 1996 the team deployed a much-needed online registration process for new NLM users. By the time NLM announced free Web-based searching of its on-line databases a year later, more than 19,000 new MEDLARS users from 96 countries had used the IGM online system to register for new NLM accounts over the Internet. Users holding online accounts could make searches with Internet Grateful Med, then take advantage of hardcopy document delivery capabilities through the Loansome Doc server built by the IGM team.

Tens of thousands of physicians have been introduced to Internet Grateful Med through outreach programs of the National Network of Libraries of Medicine and professional specialty conferences such as the annual conferences of the Radiological Society of North American, the American College of Physicians, and the American Medical Informatics Association. Instructional courses titled "Introduction to Web-Based Searching for the Librarian and Information Specialist: Using PubMed and IGM to Search NLM's Databases" and "Making the Transition: Converting to PubMed and IGM to Search NLM's Databases" were introduced to provide training opportunities for the growing user community. The announcement of free Web-based searching resulted in a profound change in NLM's online user community. As searching doubled, tripled, then multiplied by a factor of more than ten, surveys showed major elements of the user community to be 36 percent researchers, 34 percent health care practitioners, and 30 percent general public—consumers, patients, parents and other patient advocates. To handle the increasing user load, the Internet Grateful Med team developed and put in place a "server farm" system architecture for IGM. This system runs on eight small Sun Ultra servers connected through a high-speed ATM (Asynchronous Transfer Mode) network. Individual servers can be added as user load increases, or taken offline for servicing. A personal journal subset function patterned after that in Grateful Med for Windows was added, as was the capability to search by journal with assistance in finding accurate journal names from the list of more than 3,800 journals represented in MEDLINE. An incoming connection to IGM from the Health Services Technology Assessment Texts (HSTAT) enables the HSTAT user who has found a journal citation of interest to go directly to Internet Grateful Med to view the complete MEDLINE record of that article. Finally, the ability to invoke NIH Clinical Alerts from the front screen of Internet Grateful Med was added in late 1997.

By the end of Fiscal Year 1998, usage of the Internet Grateful Med system was running at more than 4.9 million World Wide Web connection requests per week. IGM users from dozens of countries made more than 800,000 searches in MEDLINE and other NLM databases during September 1998. A hardware upgrade more than doubled the capacity of the IGM server farm in preparation for a new version of Internet Grateful

Med. IGM V2.6, a major release in early September 1998, fulfilled the promise of the Internet Grateful Med gateway architecture by moving its MEDLINE searching to the PubMed retrieval engine while continuing to search 14 other databases in NLM's ELHILL mainframe system. Access to five new ELHILL databases (TOXLINE, ChemID, BIOETHICSLINE, POPLINE, and SPACELINE) was added. As of late September 1998 TOXLINE had already become the third most heavily accessed database through IGM, after MEDLINE and HealthSTAR. In the coming year, access to NLM's new TOXNET retrieval system and a new full text retrieval system will be explored. Intelligent software agents will be considered to facilitate searching in multiple databases from a single user query.

Object Oriented Information Retrieval (OOIR)

The primary focus of the OOIR project for the last year has been the move to an object-oriented database system for data storage and management. Several major milestones have been achieved in moving toward a fully object-oriented distributed system and in support of the Health Services Technology Assessment Text (HSTAT). The HSTAT collection has grown from 120 documents in 8 collections to 136 documents in 9 collections. A development area on a secondary server with limited access has been established for prototypes such as the collection of Health Services Research Data Sets used in a Library Associate's project. HSTAT usage has averaged 1,244,477 hits per month for the last year with an average of 71,443 unique hosts per month. The average number of queries per day is 3,830. Specific file/document access information is provided to the agencies submitting data on a monthly basis. The major addition to HSTAT in the past year was the ability to include a search of MEDLINE via PubMed when searching the entire collection. This is a bidirectional link between HSTAT and MEDLINE, since PubMed provides access to any document in HSTAT for which a MEDLINE citation is retrieved in PubMed. In the development version, a feature to add variants of search terms has been implemented using the SPECIALIST lexical programs. Though not yet in the operational version, this will be included in the next release of the system.

Other features currently being tested include a batch client to allow researchers to submit multiple queries from a file using either e-mail or a Web-based front end. Concept mapping between HSTAT data and submitted queries is also being tested, again using tools developed in the UMLS project. A doctoral student will be analyzing the effects of such mapping to assist the OOIR group in determining at what level of a document this might be most effective and helpful. An annotation facility is being worked on to allow developers of guideline information to discuss and comment through HSTAT as part of the guideline creation process. This capability will be implemented in the development version of HSTAT to help the Centers for Disease Control and Prevention in their efforts to produce the Guideline to Community Preventive Services. This will be the first active use of the annotation capability.

The OOIR group in the coming year will continue to focus on testing and experimenting with object-oriented database support, improved indexing methods for full-text in this environment, improved user navigation through large and complex documents, and appropriate search aids such as concepts and vocabulary assistance. Agent technology will also be investigated both as a direct aid to users and as a management tool while links between the HSTAT project and the National Guidelines Clearinghouse, the Veterans Administration, the Minnesota State Health Department, and the Centers for Disease Control and Prevention are forged and activated.

Image Processing Research

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 no cost agreement with the NLM. They are being distributed to licensees over the Internet and on DAT tape. The data sets are being applied to a wide range of educational, diagnostic, treatment planning, virtual reality, artistic, mathematical and industrial uses by over 1000 licensees in 27 countries. The Visible Human Project has been featured in well over 800 newspaper articles, news and science magazines, and radio and TV programs worldwide. 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 provided by the Visible Human data sets are 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.

The Visible Human male data set was introduced as a national resource in January 1995 when online access for the visible male data set was provided to approved licensed users. A year later in January 1996 the visible female was added to the collection. Since its availability the data set has been accessed by users within the United States and more than 30 countries. The Visible Human 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. Online demand for the 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. 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.

Availability of the Visible Human data set provides the medical imaging community a rich collection of full body anatomical images. However, its delivery highlights a number of issues related to effective methods for sharing large anatomical image files. These issues relate to the format in which the images are stored and accessed. Presently the images are stored as body cross section slices in raw files, and delivered via FTP in a non-anatomical format. This requires users to have prior knowledge of the sliced ordered sequence if they wish to retrieve a specific anatomical structure. To address this issue we have begun a second phase of the Visible Human Project to address the anatomical labeling, storage and retrieval of the data set. This extended phase includes development of a prototype anatomical management system consisting of an input data management processor, database, retrieval system, and a visualization module. Collaboration with the UMLS development team provides access to the UMLS lexicon, and the UMLS Knowledge Source Server.

During 1998 project staff have met several times with colleagues from the National Institute for Dental and Craniofacial Research (NIDCR). We have jointly sponsored a workshop on the feasibility of creating a multimedia head and neck atlas based on Visible Human data. Based on the recommendations of that workshop, NLM and NIDCR will join with the National Institute for Neurological Diseases and Stroke and the National Eye Institute to sponsor several parallel efforts beginning in 1999. One project will develop a Visible Human Head and Neck Atlas that would reside on computers housed at NLM and made available via the Internet. The atlas will have labels, voice pronunciation of anatomical terms, and other visual and explanatory information. A modular approach will be employed so that new advances in technology and segmentation techniques can be rapidly inserted as the project develops. We envision that the atlas will be immediately upgraded as new enhancements to various components are developed and added. The availability of the modular data set underlying the atlas will be enhanced by new authoring tools and end user software developed by the private sector.

Other research will explore questions concerning how to overcome deficiencies in the current data caused by the anatomical data acquisition method that was used. In general, finer resolution of all the organs and structures is desired. Specific research problems include finding new ways to prevent brain herniation as an artifact of freezing the specimens; improving imaging of eyes and their related structures; establishing appropriate techniques for arterial injection to maintain appropriate contrast; developing an appropriate technique for staining nerves on cryosections; and imaging nerves and other structures at a very fine level of detail. Computational questions of segmenting and aligning the data will also be investigated. The questions will focus on developing new computational methods for the use of composite data within Visible Human data sets, and utilizing areas of greater resolution, special staining, motion or other special explanatory information. Both sets of research projects will be undertaken in anticipation of constructing the new high resolution Visible Human Head and Neck Atlas.

A revised and final draft script for the new Visible Humans videotape was completed in August 1998. The updated videotape, which will be about 30 minutes, is titled *The Visible Humans—The Journey Continues*. The tape will include elements of the original 1996 production but will include new examples of applications of the datasets developed by some of the 1000 licensees around the world. Included in the new version will be examples from the University of Maryland School of Dentistry (muscle dural connection in the neck), the Mayo Clinic (use of 3 dimensional imaging in brain

surgery), the University of New York at Stony Brook (3D colonoscopy), the University of Colorado (celia plexis block simulation), and HT Medical (bronchioscopic simulation).

Animations of 3D rendered heart, trachea, and lungs of the Visible Human male from the 70mm film data were created based on the segmentation provided by Engineering Animation, Inc. Some of them were created using the true RGB colors of the actual data using new rendering tools, and some were created in pseudo colors using ANALYZE multiple object rendering tools. All animations can be displayed in ANALYZE.

Meetings at the Yale University School of Medicine were arranged by Dr. Carl Jaffe of Yale University's Center for Advanced Instructional Media (CAIM) in order to define the content material for a cranial nerve project. This is a Web-based teaching tool for use in an anatomy course for medical students at Yale. Areas of interest include 3D representations of the sigmoid sinus, jugular vein, mastoid air cells, internal acoustic meatus, the spatial relationships of the middle ear, and the anatomic course of the facial nerve. These selected areas are very difficult for students to visualize.

The Visible Human female CT images were used to create a 3D skull as an anatomic reference showing the location of the inner ear and its anatomy. A 3D motion sequence of the external anatomy as well as the base of the skull was created and an illustration of the cochlea and the semicircular canals were added to show their location in relation to other structures such as the internal acoustic meatus. An RGB color slice from the VHF data set that shows the vestibulocochlear nerves was then superimposed to show the correlation with the CT images of the bone anatomy. These images were integrated in a QuickTime movie which was sent to Yale for review by medical school faculty. This collaboration in which Center staff produce the 3D anatomic images that require significant computing power and Yale collaborators provide expertise in integrating them into their teaching modules, provides valuable feedback from faculty and students.

Image Storage and Transmission Optimization (ISTO)

The objective of this project is to investigate 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 Project amounts to 55GB, and when the 70mm photographs are scanned at much higher resolution, that will total about 235GB. These datasets will strain both storage and transmission resources, and ISTO was proposed as an investigation of both compression and advanced communication techniques to alleviate these problems. Research has been done toward the development of prototype lossy and lossless compression techniques. The eventual goal is to design a system combining both techniques so that storage is achieved losslessly, and data to be delivered to a user is compressed lossily, at a quality level (and therefore compression ratio) required by the user. This user-defined quality level will define the wavelet/quantization parameters suitable for lossily compressing the data prior to transmission. Specific current goals are to analyze alternative lossy and lossless compression methods for the Visible Human image data, design and develop two alternative modules, and conduct a comparative evaluation; and to evaluate alternative transmission techniques for the Visible Human image data using NLM multi socket technique and RFC 1323 over the current Internet and ATM.

To evaluate the comparative performance of compression techniques, a sample of the Visible Human image set was compressed both lossily and losslessly using public domain software for wavelet transform, DCT (lossy JPEG) and lossless JPEG. In addition to compression resulting from eliminating the background pixels in the images (over 3.5), we hypothesized that for lossless compression pixels in adjacent slices would be highly redundant and therefore promising for compression. We conducted studies of interframe redundancy using the RGB color images but also with the images transformed to the YIQ and HSV planes. Our studies showed that interframe redundancy is low irrespective of color plane. For lossy compression, we have begun studies toward the selection of wavelet transform suitable for VH images. The statistical properties of the coefficients in wavelet transformed subimages of selected VH slices were investigated using filter taps from 4 to 20 to establish the optimality of specific wavelets for this class of images. A tentative conclusion is that both 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. In 1998, experiments were conducted using the EZW technique that incorporates wavelet transform and scalar quantization. Researchers achieved compression ratios ranging from 1.8 to 2.1 depending on the slice. Also depending on the image, the time to compress ranged from 36 s to 42 s after the code was optimized to eliminate function calls and therefore stack operations. The original code for EZW was designed to compress single color-planar images, but this was modified to code three color planes with width and height an integer power of two, a reason for cropping the 2048 x 1216 images to 2048 x 1024 retaining all of the anatomy and

discarding background. The decompression time was measured to average 5.5 minutes. For lossless compression, an Adaptive Arithmetic Coding based module was develop to compress the cropped images.

Using the lossless and lossy techniques as elements, a prototype interactive access tool for the VH images was created consisting of two modules: the Image Archive Retrieval Application (IARA) and the Image Viewer Application (IVA). The IARA establishes an Internet connection to the server, searches for the desired images selected by the user using a simplified body map, responds to a quality factor chosen by the user, downloads images from the server, and stores the compressed images in the user's computer. The IVA lets the user specify a path for the compressed images, calls the inverse EZW algorithm (which is coded as a Java Native Function), lets this function read the file and perform the decompression, and displays the image for viewing. The IVA lets the user scroll through the image.

While Java was used to implement the client-server application, the source codes for the Arithmetic Coding and the EZW algorithms were in ANSI C. These source codes were not rewritten in Java, but were called and run from Java, since C programs are faster than Java programs. The Java Native Interface (JNI) provides the functions needed to call native functions from Java. The main program of the C code is written therefore as a Java Native Function and compiled as a shared library. The Java program loads this library, declares the native method, and calls it. The Java Native Function receives and returns Java arguments that are converted to C variables and vice versa. The Java applications use JNI to run the Adaptive Arithmetic coding and decoding and the EZW decompression C codes. A key point in this design is that the user specifies the region of interest as well as the quality desired. Four check boxes in the GUI lets the user select four levels of quality, the highest at CR of 20 and the lowest at a compression ratio of 100. The tradeoff is the speed at which the images arrive at the desktop and get displayed, speed being inverse to quality.

In addition to the compression work, ISTO will also continue to evaluate file transfer performance of NLM multi socket technique vs. RFC 1323 over the Internet and ATM networks, and evaluate client-server database application performance of NLM multi socket technique vs. RFC 1323 over the Internet and ATM networks. Center staff have demonstrated ATM transmission of VH images to various locations, and have built the local infrastructure (including routers and ATM switches) to connect to vBNS at OC-3 (155 Mbps) speed as a first step to participating in the Next Generation Internet and Internet-2 initiatives.

Automating Data Entry into MEDLINE

The Medical Article Record System (MARS) is a project whose goal is to develop systems that achieve a degree of automation in the entry of citation and abstract data from medical journal articles for the MEDLINE database. This data entry has traditionally been done by manual keyboarding. In response to a crisis in early 1996 in this data entry, the MARS team developed the first generation of this system that combines 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 abstracts which, if keyboarded, proves very labor-intensive. The first generation system, MARS-1, consists of about two dozen workstations of three types: scanner, reconciling, and keyboarded citation entry. In addition, the system requires three servers: a network file server, an OCR server and one to match double keyboarded citations. All workstations, networked via a LAN, are installed and operating at the work site. The current volume of production is over 600 completed records a day compared to 65 records per day when the system was installed in October 1996. Since achieving the goal of 600 completed records a day in February 1998 (one-third of the Library's total record production), the MARS system has consistently maintained or slightly exceeded this level of production.

This increase in the production volume was achieved mainly by scaling up the system to double the number of workstations, but gradual improvements were also helpful. These incremental improvements include modifying the scanning workstations to allow the operators to enter journal issue identification numbers by barcode readers instead of keyboard entry; providing on-screen tables containing Greek and special symbols to allow the reconciling operators to click from a list in the table without having to type in the words; providing a highlight-and-click capability to capitalize or de-capitalize words or entire text lines; changing icons, error messages and buttons for clarity; including the capability of entering diacritical marks directly by keyboard operators; and developing a spell check module with lexicons and ad hoc rule sets to unhighlight correct words (from the OCR) while retaining highlights for the incorrectly recognized words, and thereby reducing the burden on the reconcile operators.

The spell-check module was developed as a result of research into the use of the UMLS Metathesaurus and the Specialist Lexicon, coupled with heuristic rules related to word lengths, to increase the confidence levels of the characters in the OCR output, and thereby reduce the number of

incorrectly highlighted (low confidence) words seen by the reconcile operator. The spell check module reduced the highlighted words by about 50 percent, and yielded a 4 percent increase in the overall production rate.

In addition to these incremental improvements, a tool called *Tracker* was developed to provide management control of the production process. Tracker is a Visual Basic based system to generate chart reports reflecting MARS production rates, daily, weekly or monthly, for any time period of interest. Tracker may be used to give a bar chart of daily production levels during a certain time period using MARS "uploaded" files for the days in that period. Tracker uploads completed record data from the MARS file server into a Microsoft Access database, calculates the required statistics, and invokes a chart display (bar, pie or graph, selectable by the user) showing the numbers.

While MARS-1 is in routine production, the team is conducting R&D toward more comprehensive automation. The design of a database-centered and database-driven next generation MARS-2 with the definition of over 130 tasks has begun. The initial wave of tasks is focused on the SQL Server database design, and they include: 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 engine; and a new GUI design for the scan workstation incorporating a magnifying glass feature, a dialog box for confirming MRI, and a continue button for abstracts that span two pages; a 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.

This research toward more comprehensive automation also includes automated field entry by page segmentation (autozoning and automatic field identification) software, and the development of algorithms for reformatting field syntax. This will automate, for example, the reformatting of the author field "John F. Smith" as recognized by the OCR system to "Smith JF" as required by the MEDLINE convention. For the research in automated field identification, i.e., the labeling of "zones" from the scanned page as "author," "article title," etc., a tool called ZoneChecker was developed to aid the data collection from the scanned and zoned page images. This tool is being used to build a test/training set of labels (title, author, affiliation, abstract) to develop the automated field identification system using decision- making systems such as neural networks. Early testing of this data with the machine learning COEV tool has indicated that the neural nets as well as the decision tree generator (a public domain system C 4.5) in COEV may be promising decision-making modules. However, this testing will continue with more extensive data to also take into account real world problems of splitting and merging zones which are sources of error.

Voice recognition techniques were investigated to ease operator burden in entering certain kinds of data and system commands, as an option to keying or mouse operation. Two commercial speech recognition products, one a continuous and the other a discrete speech system, were evaluated for their capability in recognizing long alphanumeric strings (e.g., NIH grant numbers and Databank Accession numbers) as well as commands for controlling workstations. Tests conducted with 40 participants confirmed that continuous speech recognition may have a slight advantage over discrete speech recognition for dictating alphanumeric strings, although accuracy is only marginally adequate for production use. But for commands, it was found that either system works well for a limited set of commands. Macros were developed to operate the frequently used scan commands by voice, and voice command testing with both the Ricoh and Fujitsu scanners has begun.

DocView

The DocView Project has the goal of conducting R&D in advanced technologies for document delivery over the Internet. It involves design, development and evaluation of prototype client and server systems to enable access to, and delivery of, biomedical documents over the Internet, including multimedia data types. The three key elements of this project are that it addresses the NLM mission of providing document delivery to end users and libraries, uses the Internet and incorporates digital imaging techniques. The documents may either be in servers or scanned and sent over the Internet by Ariel systems, widely used for interlibrary services in the library world. While libraries and document suppliers use Ariel routinely to send documents via Internet to similar organizations, there are few options for end users to receive them directly. The DocView project fills that niche with the release of DocView 1.0 client software. In August 1998, during a visit to NLM, the Ariel product manager at Research Libraries Group, Mountain View, CA discussed the next developmental steps for Ariel and applauded the complementary properties of DocView in providing end users a way to receive documents sent by Ariel systems at libraries and document suppliers.

The DocView 1.0 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 World Wide 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.

Following extensive beta testing for two years, DocView 1.0 was released without fanfare in January 1998. By September 1998, this software was downloaded by over 1,300 new users in 68 countries plus Guam. In some countries DocView is finding exceptionally heavy use: in Germany, for instance, there are users in 90 cities. While it is not surprising that there are large numbers of interested users in industrial countries, DocView is also finding users in less expected places such as all over the South Pacific: Guam, New Caledonia, Fiji Islands and the Northern Marianas. In addition to friends or colleagues, people are finding out about DocView through AOL Netfind, Alta Vista, Yahoo and Excite which provide references to DocView's Web pages. There is no indication of difficulties in either installation or usage as reported on the DocView Email List Server which enables subscribers to post help messages to other users. NLM is not providing any support for installation or use. The major users appear to be scientists receiving documents at their desktops from Ariel workstations at their libraries.

In light of NIH's special interest in malaria research in Africa, in August 1998 a document delivery trial was done for a malaria research site in Kalifi, Kenya, 30 miles north of Mombasa. This site is run by the Kenya Medical Research Institute and Welcome Trust (UK). Using DocView, the Kenyans received two articles from biomedical journals sent from an Ariel machine at NLM. According to onsite researchers, the articles arrived in "perfect condition."

The extensive beta testing DocView underwent until December 1997 with over 60 institutions in 12 countries was invaluable in refining the software for its January 1998 release. Institutions participating in the beta tests included biomedical libraries, university libraries, international agencies, and document delivery service providers. The purpose of the beta test was to establish the technical feasibility of easy and intuitive document delivery over the Internet. Investigations focused on: performance of rapid delivery of document images to users over the Internet, design issues related to the selection of communication protocols and local archiving techniques, and the performance of user functions at the receiving workstation. The beta test involved 90 individuals each of whom ran DocView for 45 days (amounting to 11 person-years of use). The study showed that a majority of users felt that DocView has improved the way they receive library documents; they want to continue having documents delivered directly to their desktops; they would recommend DocView to their colleagues; and that DocView is easy to learn to use.

To continue further research into document delivery techniques building on DocView's success, four new interrelated subprojects in this area have begun. The first is multimode delivery of multimedia information. Based on the DocView software this system (meant primarily for document senders such as libraries) will deliver documents on demand: in addition to scanned monochrome document images it will handle grayscale and color images, audio and video; it will use the Web as a delivery mode in addition to Ariel FTP and MIME email; and it will be scalable so that multiple machines can be used if the delivery load increases. The second subproject is platform-independent Java-assisted document delivery (MedJava). This is a Java applet based on DocView software that can be delivered through the Web along with a document. The applet will provide document rendering and usage. The third subproject is a 32-bit version of DocView providing new functionalities such as file conversion (e.g., from TIFF to PDF), image editing and handling of color and grayscale images, and a personal Web document library that a user can make available to colleagues/students. The 32-bit version will be more reliable, will run on any 32-bit Windows platform, and will have its image processing algorithms tuned to run faster than in the current DocView. The fourth subproject is a document conversion server which will allow users to upload image files for conversion to alternative formats.

Workstation for Interlibrary Loan (WILL)

The WILL prototype system developed to deliver bitmapped document images on demand from libraries was transferred from the lab to a private company under a Cooperative Research and Development Agreement (CRADA) following a year long beta test at a large biomedical library. The CRADA agreement, a first for NLM, was approved by an NIH committee responsible for overseeing technology transfer to the private sector. Under the CRADA, the private company will conduct R&D required to enhance the existing WILL prototype with regard to the internal database, develop a GUI based on Visual Basic, incorporate faster scanning methods, and APIs to request routing systems other than DOCLINE (NLM's request routing system used by medical libraries nationwide). The company will

also take the steps required to make WILL a viable, practical system for widespread use in biomedical document delivery. In parallel, the research team provided the company with in house developed source code and documentation, and complemented WILL's transmission capability by developing end user software for receiving documents sent by WILL. This software, first modified to 32-bit code, included modules that transmit bitmapped document images via FTP to Ariel systems or to NLM's DocView.

The hardware required in WILL consists of a Pentium-class computer with internal fax, Ethernet and image processing boards. The software consists of C, C++ and assembly language modules, and runs under the Windows operating system. The only activity WILL requires of an operator is to scan the requested documents. It eliminates other operator decisions on recipient and delivery mode. WILL automatically retrieves interlibrary loan requests from DOCLINE, parses the requests into data fields, provides a GUI for the operator to scan requested documents, and automatically delivers document images by any of the three conventional delivery modes (Internet, fax and print for postal mail) as requested, and updates DOCLINE with status. In addition, a database in WILL retains operational information for administrative purposes. The only process that is not automated is the physical handling and scanning of the documents.

In May 1998, our CRADA collaborator demonstrated their first implementation of WILL based on the prototype software to verify that the modules work on their hardware configuration. They have started a new design using more up to date commercial software, and also eliminating the Kofax hardware that the original design requires. The elimination of this hardware board will reduce the cost of WILL to the users. In June the company's marketing personnel attended the ALA conference and exhibits in Washington DC as part of their investigation of document delivery opportunities. WILL is seen as a contribution to the nationwide interlibrary loan service by helping the U.S. medical library community meet their document delivery responsibilities, since increasing labor costs and new delivery modes are placing an increasing burden on the traditional photocopying/mailing method coupled with manual activities in request retrieval, status updating and statistics gathering.

Biomedical Multimedia Database R&D

In past issues of *Programs and Services*, a three-agency joint project named DXPNET (Digital X-ray Prototype Network) was described. This project was begun to serve as a vehicle 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. In this project, a collaboration with the National Center for Health Statistics (NCHS) and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS), Center staff took on the role of technical manager and developer. The x-rays, consisting of about 17,000 cervical and lumbar spine films, were collected during the second National Health and Nutrition Examination Survey (NHANES II), one of a series of nationwide surveys conducted by NCHS designed to provide a snapshot of the nation's health. As films they are relatively inaccessible, a major motivation for digitizing them. All 17,000 films were digitized with a Lumisys laser scanner resulting in images of approximate size 5 MB (cervical) and 10 MB (lumbar). These were quality checked and archived in our 144-platter optical jukebox.

DXPNET involved the design, development, and evaluation of prototype systems which serve as testbeds to investigate image compression techniques, especially high-yield lossy methods, and tools to interactively select compression parameters; techniques to organize images and associated textual data for ready retrieval and use; procedures and algorithms to implement transparent hierarchical storage using heterogeneous storage systems and media to match usage patterns; and multisocket transmission methods to segment large images and to send the pieces concurrently over multiple socket pairs to overcome the inefficiencies of conventional transmission protocols. A goal of the project was to determine the essential design characteristics required in systems that provide remote access to such a medical image collection, and to design and develop systems that satisfy these requirements. The systems developed include a Standardized Readings Workstation, the tool for radiologists to retrieve the images over the Internet and to enter their readings; an Electronic X-ray Archive consisting of an optical disk jukebox coupled with a RAID system; and Quality Control Workstations to enable NCHS technicians and radiologists to perform quality checks of the scanned x-rays.

With the development of these systems and the archiving of the 17,000 spine x-rays from NHANES II, the original objectives of DXPNET were met, but the digital capture of an additional 10,000 hands and knees x-ray images from the third survey (NHANES III) is proceeding. A key issue here was the scan density (spatial resolution) required to read this class of x-rays when digitized. A study to determine the required scan density was conducted by two radiologists each using a Sparc 20-based workstation developed in-house specifically for this procedure. The workstations run on Java software, the dialog framework built entirely in Java's Abstract Windowing Toolkit, and a C++ image display

module. About 200 hand and wrist images from a population of Pima Indians were used to determine the best scan density to detect bone erosions as a way to establish resolution requirements for the large NHANES III collection. The Pima x-rays were digitized at three resolution levels: 50, 100 and 150 microns. The study concluded that the 100 micron level would provided the resolution needed, and has provided the information required to proceed with the digitization of the hand x-rays from NHANES III.

In line with an interagency agreement with NCHS, quality control workstation software and associated equipment (Megascan monitor, external hard disk and a magneto-optical drive) have been transferred to NCHS to enable them to conduct quality control for the NHANES III images now being digitized. In-house engineers completed the development of this workstation, reusing a large part of the source code developed earlier for the workstation used for the image resolution study for hand x-rays. They also developed a new GUI for the workstation.

As a follow-on to DXPNET which focused on the x-ray images specifically, it was of interest to provide Internet access to both the x-rays as well as to the vast amounts of textual ("collateral") data gathered in the same NHANES surveys. For this purpose, a prototype system named MIRS (Medical Information Retrieval System), was designed to enable general access via Internet to the mixed text/image database consisting of NHANES II collateral data and the x-ray images. MIRS was developed on a Sun Sparc 10 platform and uses X-Windows and TAE Plus for interface development. Access to the data is by SQL query, and the data is organized in an Illustra DBMS. MIRS was demonstrated widely to show querying and access to the integrated database consisting of the images, the standardized readings, and the corresponding NHANES collateral data (demographic, blood chemistry lab data, medical questionnaire data, etc.).

While MIRS has been shown at various conferences to be a successful tool to gain access to the NHANES text and image data, for widespread acceptance a platform-independent Web- based MIRS system using the Sun Sparc-based system design as a foundation is being designed. The new WebMIRS design, written mostly in Java, provides input field validation; simple query formulating capability to automatically generate SQL queries; multithreaded image download and display capability enabling multiple images to be retrieved and displayed simultaneously; and support for Web-based database access using Java Database Connectivity (JDBC). The current phase of WebMIRS development is to evolve it as a tool, not just for querying the text, retrieving the text and image results, and displaying the results, but also for the analysis of the data for research purposes by allowing results to be imported into appropriate statistical packages. A significant step in WebMIRS development as a tool for Web-based analysis of key epidemiological data was taken with the first alpha test conducted by a professor of statistics at Emory University and an expert on the NHANES data. She used the system over three days of testing and delivered a detailed evaluation report. This was the first use of the system by someone with in depth expertise in analyzing the NHANES data. The test plan called for the use of WebMIRS to extract NHANES data and to export this to commonly used statistical analysis packages: viz., SAS and SUDAAN. Researchers often use SAS for data sorting and other manipulation functions, prior to handing over the data to SUDAAN for the analysis of complex correlated datasets such as NHANES.

Design shortcomings in WebMIRS found during the alpha test are being addressed. They include: one, SUDAAN's need for more data than WebMIRS can currently export efficiently; two, the absence of key "survey design variables" and data weighting variables required by SUDAAN. These are being corrected, in addition to other changes: displaying image identification numbers, improving the response time of the system, enabling the automatic display of the images at the proper heights, fine tuning the caching mechanism for variables at the client side to minimize the number of SQL requests issued (greatly speeding up the scrolling of the data explorers or tree views), modifying Image Canvas to display text information as well as to highlight the borders of a displayed x-ray image, and enabling the user to specify the preferred look and feel of the GUI, preferred results format, displayed image type, and the size of the image batch. Also, the Save Results feature which allows the user to save in a file all results retrieved from the back end server has been completed, and the results may be saved either as numeric codes (necessary for analysis) or as text description (for comprehension).

While it is still evolving, the WebMIRS system is now being used to provide images for research into vertebral morphometry at the University of Washington, and for other research purposes at the University of California San Francisco and General Electric.

Advanced Medical Imaging Tools

Consistent with the increasing trend for medical information data banks to incorporate images, tools must be available to enable users to easily search and retrieve such data over the Internet, and to evaluate the returned image data against "gold standard" or reference images. The objective of this project is to address fundamental questions in the

design of such tools that are required in the handling, organization, storage, access to, and transmission of very large electronic files in general and digitized x-rays in particular. This project arose out of design activities in the DXPNET project described above. Our current focus is the building of two tools, one as a medical image reference aid, and the other a multimedia database access tool.

One tool is a platform-independent digital radiological atlas of the cervical and lumbar spine building on prior work involving digitized spine x-rays from the NHANES II survey. The approach is to use Java software technology to create the atlas whose images, initially obtained from the NHANES image data set, will be displayable on conventional monitors as well as high resolution Megascan monitors. This tool would find application in PACS systems and in medical teaching and research environments.

As described in a previous section, a second tool is an advanced Web-enabled medical image database tool for searching and retrieving contents of biomedical databases containing both text and images, building on the Internet-accessible, client-server MIRS system developed in the DXPNET project. Techniques incorporated in this development include CGI scripts, C++ classes, the Informix DBMS and Virage datablades. Java applets are used for client enhancements. Using this platform-independent tool, access to the text and images in the NHANES databases will be demonstrated and evaluated. These databases, based on the Informix DBMS, will contain NHANES II information relating to the national prevalence of osteoarthritis, both textual information and cervical and lumbar spine x-ray images are to be included in this database; NHANES III information related to osteoarthritis, osteoporosis, functional health of the elderly, and other NHANES focus areas of interest as defined by NIH researchers; both textual information as well as hand and knee x-ray images are to be included in this database.

Research is also proceeding toward a "query by image content" capability to efficiently exploit the rich visual information in large biomedical multimedia data banks. This capability is expected to be critical for efficient information extraction from biomedical image data in the future. Medical information applications for this tool include delivering 3D imagery across the Web from the Visible Human image set and retrieval of digital x-ray images from the NHANES II and III collections. This research is proceeding jointly with radiologists and engineers at Georgetown University. The first step is the segmentation of vertebrae in the spine x-ray images from NHANES II. About 300 images were selected by first searching the WebMIRS databases for subjects who meet particular criteria for age and back pain symptoms, doing a batch search for the required image ID numbers, identifying the jukebox platters on which the images reside, and reading the images to hard disk after which they are recorded on 8mm tape. A radiologist specialized in musculoskeletal x-rays uses a system to display and record segmentation information from these images. When this work is complete, the data will be entered into the WebMIRS databases and will be searchable by users.

Education and Training

The Learning Center for Interactive Technology

The Learning Center for Interactive Technology offers a wide range of computer-based information and educational technologies representing stand-alone, intranet, Internet and World Wide Web applications in the health sciences. In FY 1998 979 individuals representing over 50 distinct groups visited the Learning Center, bringing the grand total of visitors to just under 12,000. In addition to the traditional constituency of health professional teachers, students, administrators, and researchers, the Learning Center continued experiencing increasing numbers of visiting high school and college students. This increase has been due in large part to the great interest generated at the high school and college level by the Visible Human Project and the availability of a variety of Visible Human applications in the Learning Center. Another contributing factor has been the inclusion of the Center in many of the general tours provided for visitors to the National Institutes of Health.

In addition to the enormous interest in multimedia applications of the Visible Human Project, visitors were also interested in Internet and intranet-based information and educational resources in medicine; World Wide Web based instructional resources, particularly problem-based tutorials and simulations; authoring tools for developing and evaluating network accessible instructional resources; creating distributed learning environments in support of health professions education; and using online collaboration tools.

Individuals and groups of visitors used the Learning Center to learn about and access a wide variety of biomedical and educational resources available on the Internet and World Wide Web. Staff provided demonstrations and facilitated hands-on practice with new and emerging online technologies and resources and assisted visitors in reviewing existing Internet accessible multimedia resources in medicine, using the Learning Center's Web server as a focal point (http://tlc.nlm.nih.gov/).

The Center re-opened in January 1998 after closing for a major renovation in December, 1997.

The modifications made the physical environment more amenable to conducting demonstrations and more formal workshops, tutorials, and seminars. Three workshops were conducted on how Internet, intranet, and related communication technologies can be used for effective learning. Fourteen tutorials on these topics also were conducted in collaboration with the Radiological Society of North America. Center resources were used by summer students and teachers from NLM's adopted school, Coolidge High School, to develop materials and prepare their project presentations. The Center also inaugurated its own Web server, moving some resources residing on the Cognitive Science Branch server and developing new ones. Tutorials developed for workshops have been placed online. A companion server for streaming video in some of the tutorials was also put into service. A program is under development that will enable visitors to schedule appointments online.

Distributed Learning Environments

The general research and demonstration focus of the distributed learning environments project is to investigate innovative means for assisting health science institutions in their use of online distance learning technologies. An important subgoal of the project is to develop, test, and utilize collaborative strategies that institutions can use to successfully integrate a wide range of technologies. The Learning Center's staff are working with the University of Missouri Health Science Center and other institutions to define the scope of the project and determine specific research and demonstration goals and milestones. Towards this end, staff at the Learning Center and at the University of Missouri have completed literature reviews of research in areas important to project implementation: adult learning; cooperative learning, problem-based learning, computer-supported collaborative work, and diffusion/adoption of innovation. Each review summarizes findings and contains a list of issues and questions for further research. The reviews are currently being compiled into a handbook for distributed learning that will be particularly relevant to institutions interested in resource sharing and collaboration. The handbook, which will be printed and also placed on the Learning Center Web server, will be used as a resource for workshops and seminars on distributed learning that will be offered onsite at the NLM and also at a distance. Staff worked with NLM's Specialized Information Systems Division and a Library Associate to assess a Web-based AIDSLINE tutorial and its video streaming components. Staff also completed reviews of selected online collaboration tools and tools for video conferencing. A multi point desktop video conferencing system using standard TCP/IP communications is currently being tested in the Learning Center for possible use in the project.

Digital Microscopy System

In 1994 project staff set about attempting to determine the parameters that govern the indexing of a library of medical photomicrographs so that eventually a library of such materials can be available to medical practitioners, researchers and students via the World Wide Web. In particular, the project intended to determine whether, as had been suggested by the Board of Scientific Counselors, such images could be described using concepts found in the UMLS Metathesaurus. A well-defined subset of the universe of such images was chosen for the tasks. In 1996 the project team installed a Web server that provides indexers with the capability of having on a single screen a photomicrograph and the input screen for the Metathesaurus. Also in 1996 a collaborative effort was initiated with the Department of Pathology at the University of Pittsburgh and WWW pages were developed that permit indexers to see images stored on the project Web server and to enter keywords and diagnoses. During 1998 indexers at Pittsburgh and NLM used the system to index photomicrographs residing on servers and NLM, Pittsburgh and the University of Illinois at Champaign-Urbana. We found the HTML interface difficult to use by indexers not thoroughly familiar with the Metathesaurus and UMLS. Since the goal is to have a Web interface that will permit pathologists to index images in a distributed database system, we set out early this year to redesign the interface to make it similar to those used by Internet Grateful Med and PubMed since most practitioners are accustomed to using those interfaces. At the end of 1998 the new interface is ready for testing and indexing of a large distributed image database can proceed. The project has also continued to investigate the possibility of converting all of the CCDS Basic Medical Pathology lessons for use on the Web. Conversion of the first lesson on Neoplasia (Neoplasia I: Benign and Malignant States) was completed in 1997 and put on the project Web server. It has been used by hundreds of students and faculty and given high ratings. In July 1998, the site was designated as an "Editors' Choice" by OncoLink at the University of Pennsylvania. The second neoplasia lesson (Neoplasia II: Metastasis and Differentiation) was converted and placed on the site in September 1998.

Cervical Cancer Project - Screening for Life

This instructional program, designed to promote the early detection of cancer, was originally produced as a videodisc, redesigned and produced in

a Compact Disc-interactive (CD-I) format, and then made available as a Web version, incorporates text, graphics, still pictures and video clips. Center staff, working with content experts from the National Cancer Institute, have now updated the prototype Web site to include new material and additional links to cancer information. The new material includes a new video clip talking about CYTYC Corporation's ThinPrep liquid based pap test method. The project now has updated mortality charts and links to NCI's Surveillance, Epidemiology and End Results (SEER) Web site as well as other "Cancer Facts" sites.

Motion Disorders Video Project

Center staff traveled to New Haven, Connecticut in May 1998 for the production of phase one of the Motion Disorders Video Database Project. In conjunction with the Motions Disorders Clinic, the Center for Advanced Instructional Media, and the Department of Biocommunications, all at the Yale University School of Medicine, two days of remote video production were completed. Fifteen patients were videotaped performing standard, diagnostic routines including hand and foot exercises, walking and reading. All of the patients recruited for this first set have been diagnosed with varying degrees of Parkinson's Syndrome. The patient set ranged in age between 40 and 75. Six of the patients were female and nine were male. In addition, some of the patients were videotaped before and after prescribed medications, to assess the effect on the individual expression of the disease.

High-end Sony digital Betacam cameras were used to guarantee the highest quality images. The videotape was edited in the AVID nonlinear editing suite, combining the two camera video into a dual screen, real-time presentation. Video of the first three patients has been forwarded to Yale for review and feedback. The second patient set, featuring non-Parkinson's patients in the Motion Disorders Clinic is scheduled for videotaping in October, 1998.

Medical Informatics Training Program

The Medical Informatics Training Program (MITP) was established in 1996 with the goal of supporting medical informatics training opportunities at the NLM. The program provides participants an opportunity to train on site at the NLM and collaborate actively in Center research in a variety of areas of medical informatics. Participants work closely with research staff, have access to on-site resources and facilities, become familiar with and make significant contributions to research, gain an understanding of the Center's programs, and have the opportunity to make professional contributions to their fields. The training programs were further developed and formalized during 1998. Program staff defined application procedures, developed formal selection procedures, established a committee of Center staff to review applications, and developed stipend guidelines. The training program was advertised in a variety of places including NLM's Web site, professional listservs, a selection of appropriate professional publications, and in a mailing to U.S. medical schools. This year, the MITP review committee reviewed 14 applicants. Of these, the Center sponsored three; three are being processed for arrival in the next few months; and three have been identified as possible candidates pending further information exchange.

This year the program sponsored 17 students ranging from high school through graduate school in its summer training program. The students worked on projects at the Center with close supervision of Center staff. The summer training program ended with a day of presentations in which each student gave a talk about his or her work. We also participated for the first time in the NIH/AISES (American Indian Science and Engineering Society) Summer Work Experience Program by sponsoring a student from University of Arizona. The student was selected from a pool of American Indian students from across the U.S. to work with us and to have an enriching experience living near the nation's capital.

The NIH Clinical Elective in Medical Informatics was again held at NLM in March and April, 1998. The Center trained eight fourth year medical students from various U.S. medical schools. The elective included lectures from outside speakers and from NLM staff. The students spent numerous productive hours working on ongoing research projects at the Center and presented their results in a day-long seminar for the NLM community.

Educational and Information Program Support

Center staff produced an 8-minute video titled, *Frankenstein: Discovering the Secrets of Nature* for presentation at the Fifty Year Reunion of the Markle Scholars in Academic Medicine, in Phoenix in September 1998. This original program features video of the exhibit, scenes from the 1931 Universal Studio's *Frankenstein*, the Visible Human Project, and footage shot in the anatomy lab at Johns Hopkins University School of Medicine. The video utilized four video and six audio tracks complete with music, sound effects, professional narration and visual effects to create a compelling synopsis of the content and theme of the exhibit. The 1998 Leiter Lecture was telecast live in May from the Lister Hill Center Auditorium via satellite. This marks the first time that NLM has originated a satellite uplink from the Lister Hill Auditorium and utilized fiber optic cable for the entire signal routing path. The video

signal was sent via fiber optic line from the auditorium to Building 10 for routing to the Washington International Teleport in Alexandria, Va. The Teleport uplinked the signal to the SBS 6 satellite.

The upcoming History of Medicine Division exhibit, on asthma titled, "Breath of Life," is scheduled to open in March 1999. Center staff is currently providing information and expertise in the identification and implementation of interactive technologies, and enhanced visual, didactic concepts as they may be incorporated into the design of the exhibit including related interactive products. In an effort to provide a new arena for interaction with children visiting the asthma exhibit, Center staff are developing a unique interactive program which is intended to provide both an engaging and instructional experience when visiting the upcoming exhibit. An existing commercial game will be enhanced with educational material about asthma. Graphics personnel developed several Website design prototypes for the Asthma Planning Committee to review. The designs incorporate specifications required by the committee, and offer unique opportunities for overall site design and navigation. It is also expected that a half hour documentary on asthma will be produced along with several short, interactive video and audio selections. There are plans to produce a "walk through" tape describing the exhibit after it has been erected.

The new Relais system method was videotaped in December 1997 and the previous method of handling Inter-Library Loans was videotaped in March 1997. The Relais videotape was completed and shown at the Board of Regents meeting in January 1998.

Center staff produced an enhanced version of a Public Service Announcement by Dr. Michael DeBakey, which had been originally videotaped in January 1997. The new version featured Dr. DeBakey's message, as well as a video of a variety of NLM outreach projects, and a medical information query via PubMed on the World Wide Web. The first installment of the public service series was launched at the HII Conference at Georgetown on April 28, 1998. The 30-second announcement is accessible online on the Center Web site.

The appearance of Rep. Newt Gingrich at the HII conference at the Georgetown University Conference Center on April 28 was videotaped and edited for the Office of Communications and Public Liaison and highlights from the tape were shown at the Board of Regents meeting. Prior to the conference, selected video clips from existing NLM video material were edited for showing at the conference.

Support for Frankenstein Exhibit

For the current exhibit, *Frankenstein: Penetrating the Secrets of Nature*, Center staff provided technical and organizational support and expertise in the video and graphic components of the exhibit, as well as production support for derivative programs stemming from the exhibit. Specifically, we coordinated facilities and equipment support for the evening Film Series, provided audiovisual support for exhibit events, and developed numerous video programs and interactive kiosks. We worked with the History of Medicine Division exhibit personnel on the production of an extensive exhibit catalogue, and the development of graphic identities for the exhibit, and related programs and promotion. Center staff produced two open-captioned videodiscs created for kiosks. One contains excerpts from four Frankenstein based Hollywood motion pictures while the other kiosk is running a 5 minute version of *The Visible Humans - A Step toward Tomorrow*. An 8-minute tape showing highlights of Boris Karloff's career and rare color footage of the monster was produced to be shown at the opening reception. The audiovisual support for the opening reception on October 30, 1997 included videotaping the entire event which was then edited to produce a 10-minute "Staff Appreciation" tape. A 9-minute videotape, featuring edited segments of an interview with Sara Karloff, a welcome and introduction by NLM Director, Donald Lindberg, and a review of the objectives of the exhibit by the Chief of the History of Medicine Division was also produced.

Office of the Public Health Service Historian

Much of the time of the staff of the office of the Public Health Service Historian in 1998 was devoted to Public Health Service Bicentennial activities. The Office was involved in the planning and implementation of the major Bicentennial events in Boston and Philadelphia. The Office also developed a 10-panel portable Bicentennial exhibit which was displayed in full or in part at various sites (including the lobby of the National Library of Medicine). The exhibit "Doctors at the Gate: the U.S. Public Health Service at Ellis Island," co-curated by the PHS Historian, was on display at the National Museum of Health and Medicine from February 5 through June 21, 1998. The Historian's Office also worked with other LHNCBC staff to prepare a revised version of the video "Plagues and Politics: The Story of the United States Public Health Service." Staff historians published several articles and delivered various lectures on PHS history and continued to answer historical queries and to cooperate with other agencies in assuring the

preservation of important documents and artifacts related to PHS history.

Resource Support and Development

System Security and Advanced Network Planning

Work during this year included the areas of system security, the NLM network and the Next Generation Internet. System security enhancements concentrated on the development of logon authentication policies for NLM and the testing of intrusion detection programs. Work continued on the development of NLM's ATM (Asynchronous Transfer Mode) network and on plans for NLM participation in the Next Generation Internet project.

Deployment of the secure shell (SSH) authentication system on Unix workstations, Intel and Macintosh platforms began in 1997 and was completed this year. The SSH system encrypts all transactions so that secret information such as passwords cannot be intercepted by hacker programs called network sniffers. In addition, the SSH system provides a means of secure access and file transfer for root accounts over the network. Experience with the SSH system has resulted in the drafting of an NLM policy on authentication.

Last year the Internet Grateful Med server farm was added to the ATM network. This year several additional servers and workstations were added to another ATM Lane network. Fast Ethernet switches have also been installed to provide another means of network access. An upgraded router will provide increased bandwidth for the Fast Ethernet local network traffic and for the Next Generation Internet connections.

Project staff have been planning the Next Generation Internet connection to the vBNS (very high speed Backbone Network Service) network. NLM was recently approved for connection to the vBNS and will be the first federal agency to do so. NLM, with an OC3c connection to the vBNS, will then have a high bandwidth connection to more than 100 universities on the vBNS. This connection will augment NLM's existing connection to the DOD supernet via the ATDnet.

Emerging Network Retrieval Protocols

The Emerging Network Retrieval Protocols group has made significant progress on a new Web services infrastructure. This includes five locally authored software elements: a context-adaptive user interface, a staff directory database having a distributed maintenance system, an automated document submission process (allowing distributed authority for content maintenance on a centralized web server), a PDF document production service, and a heavily modified commercial site map system. The NLM-developed components use the Apache Web server in conjunction with the PHP 3.0 server-side scripting language and the MySQL relational database system.

The group also has completed most of the work on a number of UNIX infrastructure improvements, most prominently the perl-based depot system for shared software repositories and the Common User Environment (CUE) version 1.0, a set of files for initializing and standardizing the sh, csh, ksh, OpenWindows, and Common Desktop Environment applications running on UNIX systems. The work on depot was done in collaboration with personnel at the Advanced Research Laboratory at the University of Texas at Austin.

An Interagency Personnel Agreement with the University of California, San Francisco for the collaboration of Mr. John Kunze will explore several examples of persistent naming mechanisms for digital documents, specifically in the context of the World Wide Web. Alternative approaches to this problem include the handle server from the Corporation for National Research Initiatives, methods based on the Domain Name System, and the Persistent URL (PURL) methodology of the Online Computer Library Center. A network-based teleconferencing capability has been set up to facilitate this cross-country collaboration.

Next Generation Internet Demonstration

A Next Generation Internet (NGI) demonstration was held in Washington DC in March 1998 at Highway 1. It was organized by the White House Office of Science and Technology Policy and the White House National Economic Council to demonstrate the power and potential of NGI. NLM was one of the many federal agencies to demonstrate application projects. NLM's Visible Human Project demonstration required 3D imaging as well as a live video source originating from NLM facilities and transmitted over a high speed network. The objective of the demonstration was to show the difference in speed between the current Internet and the future NGI. One part of the demonstration consisted of a Web-based anatomic browser that allowed the user to select a structure that could be downloaded to a local machine from a server at NLM. A second part of the demonstration included remote access of ANALYZE running on an SGI workstation at NLM. A series of large, motion sequences of 3D rendered images based on the Visible Human data set were shown to demonstrate the advantage of a high speed network. In addition, large QuickTime files of 3D rendered organs from the University of Hamburg were also accessed remotely and displayed on the local workstation.

Internet 2 Demonstration at Highway I in Washington DC

An Internet 2 demonstration, in collaboration with Yale University, was held in downtown Washington, DC, in April 1998 at Highway 1. The application demonstrated was titled "Interactive 3D Education in the Life Sciences." Computer-based 3D educational materials were shown using complementary resources residing at Yale University and the NLM. These included anatomic structures segmented on an SGI Onyx workstation at NLM and Web-based curricular materials that the Yale Center for Advanced Instructional Media team created for classroom delivered educational multimedia. The NLM and Yale used Macintosh computers on the Internet to collaborate with each other concerning real time design changes to the resources. The emphasis of the demonstration was the effectiveness of the interaction between Yale and NLM in real time facilitated by the high speed network. Various members of the Internet 2 community participated in this event including executives and engineering representatives from the member universities, congressional and White House aides, agency staffers, and the press.

Audiovisual Support Activities

Center staff continue to upgrade the equipment used to support meetings held in the Lister Hill Center Auditorium and the NLM Board of Regents Room. Center staff also provide preventive maintenance for audiovisual recording, playback and projection equipment used in other meeting rooms throughout the Library.

Engineering Laboratories

Document Imaging Laboratory

The Document Imaging Laboratory supports the DocView, WILL and MARS-2 research and design projects. Housed in this laboratory are advanced systems to electro-optically capture the digital images of documents. Subsystems are available to perform image enhancement, segmentation, compression, OCR and storage on digital optical disk media. The laboratory also includes high-end Pentium-class workstations running under Windows 95 and NT, and Novell Netware-based networks for performing document image processing. Specific equipment developed in-house includes document capture workstations for both loose-leaf and bound volume documents. Other systems have integrated optical disk drives and high resolution softcopy display devices for documents. These are configured into systems that 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.

Image Processing Laboratory

The Image Processing Laboratory supports the investigation of image processing techniques for both grayscale and color biomedical imagery at high resolution. It consists of computer and communications resources and image processing equipment to capture, process, transmit and display such high-resolution digital images. The machines include a Sun Ultra Enterprise 4000 server with magnetic disk storage capacity of 10 Gigabytes and 512 Megabytes of RAM. This machine operates as a file server for all Unix machines in the branch and as an interdepartmental resource for commercial and public domain packages. The branch also has two SPARC 10's and two Ultra 10's for development, a SPARC 20 for hosting the branch's Web server, a color scanner, and a SPARC Storage Array Model 100 RAID system. A Sun 670MP hosts an optical disk jukebox, a stand-alone magneto-optical single-density (600 MB) drive and 6 Gigabytes of magnetic disk storage; the laboratory also has a stand-alone magneto-optical quad-density (2.3 GB) drive hosted on a SPARC workstation. The medical x-ray imaging systems are hosted on a SPARC 10 with a high resolution DOME SBUS video card and a high resolution Megascan monitor. A similarly equipped SPARC 20 serves as a developmental platform for workstations to read digital x-rays.

Large-volume storage is provided by a jukebox containing 144 5-1/4 inch rewriteable magneto-optical platters, each formatted to contain files in the format required by the Unix file system. Each platter has a usable storage capacity of 586 MB, for a total jukebox usable storage capacity of 81.5 GB. In addition to the jukebox, mass storage is also provided by a Sun SPARCstorage RAID system consisting of 18 1.2 GB hard drives, six independent fast buffered SCSI-2 buses, and connected to an SBUS card in a SPARC 20 via a 25 MB/s fiber channel connector. A newly acquired pair of Sun A5000 SPARCstorage Arrays will provide 17 9.1 GB drives of magnetic storage connected to the Ultra Enterprise 4000 via a 200 Mb/s fiber channel connection. The Ultra Enterprise 4000 and 670MP 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, Token Ring 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.

External Research Support

Telemedicine

NLM is the lead agency within the Department of Health and Human Services for the Federal 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.

Nineteen telemedicine projects are currently being funded, affecting rural, inner-city, and suburban areas, with a total budget of $42 million. The 19 multi-year telemedicine projects, located in 13 states and the District of Columbia, will 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 has reviewed and will apply the recommendations from two NLM sponsored National Academy of Sciences 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*).

Next Generation Internet

A three-phase effort to support test-bed projects which demonstrate the use of NGI capabilities by the health care community has been designed. Funded projects should improve our understanding of the impact of NGI capabilities on the nation's health care, health education, and health research systems in such areas as cost, quality, usability, efficacy and security. Phase 1, which began in late 1998, is a 9-month planning effort. Each plan must identify the relevant outcomes, processes and cost variables and present a strategy for their measurement. Phase 2, a 2-year effort, will support the implementation of these plans within a limited geographic scope. Phase 3, a 2-year effort, will test the scalability of phase 2 projects to a national scope.

A study by the Computer Science and Technology Board of the National Research Council is currently being funded to define the technical capabilities that the NGI must provide in order for it to support the demands of health care applications. The study will identify likely health care applications; examine their demands for such characteristics as bandwidth, quality of service, security, and access; and recommend an appropriate strategy for implementing these capabilities within the NGI. A special effort will be made to distinguish those capabilities that are unique to health care applications from those more generally demanding of the NGI.

National Center For Biotechnology Information

David Lipman, M.D.
Director

The National Center for Biotechnology Information (NCBI) was established by Public Law 100-607 in November, 1988, as a division of the National Library of Medicine. The establishment of the NCBI reflects the importance of information science and computer technology in the understanding of the molecular processes that control health and disease. The Center has been given the responsibility to:

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

NCBI consists of senior scientists, postdoctoral fellows, and support staff working in computational biology research and the application of these methods to building public information resources. These scientists have backgrounds in medicine, molecular biology, biochemistry, genetics, biophysics, structural biology, computer and information science, and mathematics.

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 has established a network of national and international collaborations and also closely coordinates its activities with other NLM divisions.

Database Building

GenBank—The NIH Sequence Database

NCBI is responsible for all phases of GenBank production, support, and distribution. GenBank is the NIH genetic sequence database, an international database that collects all known DNA sequences, and a critical research tool in the analysis and discovery of gene function.

The growth rate of GenBank continues exponentially. In FY 1998, over 1,000,000 new sequences were added. The two-million mark in terms of number of sequences was reached in January 1998 and the three-million mark will probably be reached by December 1998. It took GenBank 16 years to accumulate the first million sequences, 16 months to acquire the second million, and will take less than a dozen for the third million. Likewise, the first billion basepairs were accumulated over a 17-year period, and the second billion basepair mark was reached in 14 months. For the coming year, the rate of growth shows no signs of abating. The continued support for sequencing of the human genome and refinements in automated sequencing technology promise that the exponential growth rate will only increase.

Complete genome sequences for *Archaeoglobus fulgidus*, *Borrelia burgdorferi*, *Aquifex aeolicus*, *Mycobacterium tuberculosis* H37Rv, *Treponema pallidum*, and *Chlamydia trachomatis* were added to the Entrez genomes database. Other organisms for which there was a 50 percent or more increase in the number of GenBank records are *Schizosaccharomyces pombe* (fission yeast), *Rattus norvegicus* (Norway rat), *Drosophilia melanogaster* (fruit fly), *Fugu rubripes* (pufferfish), *Oryza sativa* (rice), *Mus musculus* (house mouse), and *Plasmodium falciparum* (malaria parasite). Sequences from these organisms continue to provide valuable clues for understanding the functioning of human genes. There was an increase of 64 percent in human sequences, which reflects the accelerated pace due to the Human Genome Project.

NCBI produces GenBank from thousands of sequence records submitted directly from authors prior to publication; records submitted to our international collaborators, EMBL (European Molecular Biology Laboratory) and DDBJ (DNA Data Bank of Japan), are added daily. 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.

Comprehensive coverage of all sequence data, protein as well as DNA, is provided by GenBank along with the corresponding MEDLINE bibliographic information, including abstracts, and increasingly, links to publishers' full-text documents.

NLM has expanded its journal coverage to include all journals that regularly contain sequence data even if they are in nonmedical domains, e.g., applied biotechnology. GenBank is a key component in an integrated sequence database system that NCBI developed in order to serve as a single, comprehensive source of all known DNA and protein sequence information. The integrated database offers researchers the capability to perform seamless searching across all available data including the MEDLINE abstracts linked to the sequence data.

GenBank has evolved to contain several types of DNA sequences, from relatively short Expressed Sequence Tags (ESTs) to assembled genomic sequences several hundred kilobases in length. EST data obtained through cDNA sequencing are critical to understanding gene function and continue to be heavily represented in GenBank. As such, additional annotation is available for these sequences as part of the separate dbEST database. Another rapidly increasing segment of GenBank is the GSS (Genome Survey Sequences) division, which increased by over 800% in FY 1998. These sequences are similar in nature to ESTs, except for their genomic origins, and additional data on each sequence is kept in the separate dbGSS database. The newest division, High-Throughput Genomic (HTG) sequences, was created to accommodate a growing need to make "unfinished" genomic sequence data rapidly available to the scientific community. Incomplete sequences (phase 1 and 2) are updated in the HTG division as work progresses, and moved to the relevant organismal GenBank division upon completion (phase 3).

An international collaboration with the EMBL Data Library at Hinxton Hall, UK, and the DNA Database of Japan in Mishima facilitates the exchange of sequence data worldwide through a shared 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 efforts and ensure the comprehensiveness of the database.

PubMed

NCBI has continued development of PubMed, a web-based system that provides two-way links between bibliographic citations and the web servers of collaborating publishers. Using this system, complete journal articles (text and graphics) can be retrieved. As a superset of MEDLINE, PubMed contains Pre-MEDLINE, MEDLINE, and articles from selectively indexed journals that normally would not be selected for MEDLINE indexing. Retrieved citations, some of which point to journal publishers' Web servers for full-text articles, can be expanded using algorithms that find similar articles. During FY 1998, Version 2.0 was released, incorporating new search capabilities, additional search fields, a document ordering system (Loansome Doc), and a browser for the MeSH indexing vocabulary. The search software was rewritten to use bit arrays, both reducing processing time for queries with highly posted terms and making it practical for Internet Grateful Med to send queries to PubMed for processing. Data on users and usage was collected during a one-week survey via PubMed; the largest user groups were researchers and physicians, and the leading uses of the data were academic research, education, and patient care.

The "See Related Articles" feature, which expands retrieval to papers that are similar to ones identified as useful, was adapted from algorithms for "neighboring" in the DNA and protein databases. These algorithms continue to undergo research and refinement to improve retrieval. Continuous collaboration with MEDLARS Management Section and other NLM staff is used for screen redesign, addition of features, and standardization with MEDLINE protocols. Participation of the publishing community continues to increase, and the number of links to full-text journals has almost tripled in the first complete year of free PubMed availability. At the end of FY 1998, there were links to nearly 300 full-text journal sites. Some publishers provide access only by subscription, and others provide free access to the current issue, with subscription required for back issues.

PubMed usage has grown extraordinarily rapidly, with over four million different users accessing it during FY 1998. PubMed has the highest growth rate of NCBI's services, and each day, half a million searches of PubMed are performed.

Other Specialized Databases and Tools

NCBI continued to be active in the design, implementation, and distribution of new databases and software tools for the molecular biology community. Of particular interest to the scientific and academic communities, as well as to the lay public, is the release of a new version of the Gene Map of the Human Genome (GeneMap '98), which supplements a paper in the October 23, 1998 issue of *Science*. Produced in collaboration with a team of 64 scientists worldwide, the map charts the chromosomal locations of over 30,000 human genes, about twice the number of the earlier version of the map, which was released in October 1996. NCBI's UniGene (Unique Human Gene Sequence Collection) was used in the compilation of the Gene Map, and both are available on NCBI's Web site. The 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. For

scientists, the Gene Map web page contains background information on STS (sequence-tagged site) markers, RH (radiation hybrid) mapping, and provides marker and map views, as well as options to do text or position searching. The site also links to a new web page, "Genes and Disease," designed for the public and students. It includes descriptions for some 60 genetic diseases and links to databases and organizations that provide additional resources.

Another database of importance to the participants in the Human Genome Project, called dbSNP, was initiated in FY 1998. This database will contain single nucleotide polymorphisms, which exist at defined positions within genomes. dbSNP contains both the experimental conditions used to find each mutation and each mutation's observed variation for populations and individuals. The information in dbSNP will be useful for physical mapping, disease association, and surveys of population structure.

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. 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 over the Internet to a BLAST server at the NCBI. In a few seconds, the BLAST server compares the user's sequence with up to a million known sequences and determines the closest matches.

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 e-mail service has a public key encryption option to guarantee the confidentiality of user data as it traverses the public networks. The popularity of BLAST has stressed the existing computing capacity and additional computing resources are being added to accommodate the growing volume of users and expansion of the sequence databases.

To streamline the BLAST services, a new version of the program with a new query engine used by all BLAST servers was released and the older "experimental" service was retired. The BLAST web pages were redesigned for ease of use and reduction of network traffic. A graphic overview for BLAST search results was added to the Gapped-BLAST, which is the primary BLAST tool. A new service called Pattern Hit Initiated BLAST, or PHI-BLAST, was added to the BLAST suite. As input, PHI-BLAST expects a protein query sequence and a pattern contained in that sequence; it searches whatever sequence database is specified for protein sequences that contain the input pattern and have significant similarity to the query sequence in the vicinity of the pattern match.

BLAST analysis tools were introduced for subtyping an HIV-1 sequence against a set of reference sequences and for searching unfinished microbial genomes not yet in GenBank. In addition, a specialized page was added for searching the *Pseudomonas aeruginosa* database, sponsored by the Pseudomonas Genome Project, and work is nearly complete on a special service to run BLAST queries on sequences by specific organism.

Several specialized web pages were released in FY 1998. The Human Genome Sequencing Index (HGSI) was produced in collaboration with the Human Genome Organization (HUGO). The HGSI will collect and distribute up-to-date human genome sequencing information on chromosomal regions that a sequencing center is sequencing or plans to sequence. Work was completed on the High Throughput Genome (HTG) Web page, which contains information for both regular GenBank users and sequencing centers about HTG sequence submission, processing, and access. A new web page of Clusters of Orthologous Groups (COGs) was developed that compares protein sequences encoded in seven complete genomes representing five major phylogenetic lineages. A new web page on Malaria Genetics and Genomics, developed in collaboration with the National Institute of Allergy and Infectious Diseases, contains Plasmodium falciparum sequences databases, genome maps, and linkage markers.

At the close of FY 1998, in collaboration with the National Human Genome Research Institute (NHGRI), it was decided that the Online Mendelian Inheritance of Man (OMIM) database, a continuously updated catalog of human genes and genetic disorders, will be produced under contract to NCBI. NCBI developed the data maintenance software and web interface for OMIM and will continue to offer web access to OMIM.

The Cancer Genome Anatomy Project (CGAP) web page was augmented with a Digital Differential Display (DDD). DDD is a computational method for comparing sequence-based gene representation profiles among individual cDNA libraries or pools of libraries. CGAP, a collaborative effort with the National Cancer Institute, brings together data on gene expression in normal, precancerous, and malignant tissues. Design was begun for two new databases related to CGAP; SKYdb, which will contain chromosome structures in cancer, and cCAP, the Cancer Chromosome Aberration Project, an initiative supported by NCI, NHGRI, and NCBI. The Mitelman chromosomal aberration summary, which appeared in the April 15, 1997 issue of *Nature Genetics*, was made available

on the NCBI web site in conjunction with plans for cCAP.

In order to streamline production of the explosive growth in genomic sequence data, work has begun on databases that will be used to build genome records, serve as a back-end for data retrieval of genomes, and contain the official gene symbols. The latter will also serve as an intercommunication tool for staff at NCBI, HUGO, and OMIM.

The members of the NCBI taxonomy group plan the overall structure of the taxonomy database and web pages, monitor the literature, and maintain contact with off-site taxonomy advisors and the collaborating sequence databases in Europe and Japan. They add new species or perform other edits to the taxonomy database when so required by new systematic insights and guide the NCBI indexing staff on taxonomic issues.

The taxonomy group is currently composed of three biologists and one computer scientist. A fourth biologist with primary expertise in Metazoan zoology is being sought.

The taxonomy database contains 45,600 organisms represented with one or more sequences in GenBank (42,600 at the species level). About 900 species a month were added to the database in the past year, for an overall increase of 34% for the year.

The ongoing taxonomy database collaboration with the EBI and the DDBJ nucleotide sequence databases and with the NCBI structural biology group has continued to flourish over the past year. The taxonomy browser now includes explicit links to protein structures, in addition to nucleotide and protein sequence entries. The Swiss-Prot protein sequence database curators are developing procedures to use the NCBI taxonomy database nomenclature and classification.

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, and has been expanded to include the integrated set of PubMed, MMDB 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 yet intuitive way of accessing the data.

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 graphical views of sequences and chromosome maps. The structure viewer, Cn3D, permits visualization of 3-dimensional protein structures. A new feature was added to Batch Entrez that allows a search query to be used for retrieving a large set of sequences to a local disk file. Retrievals can also be based on a list of accession numbers or a specific organism.

Entrez users submit over 500,000 text searches and 70,000 sequence similarity searches daily. The web continues to grow as the principal mode of access to NCBI services, with over 95% of searches conducted via the web. Over 90,000 users access NCBI per day. Of these users, some 20,000 have accessed exclusively the molecular biology databases. Because of the mission-critical nature of NCBI’s computing platforms for PubMed, Entrez, BLAST, and other services, an extensive program in system monitoring was begun this year. Based on measurements taken every 15 minutes, the average time to load NCBI's home page is half a second, and overall availability has been better than 99 percent.

Other Network Services

Usage of NCBI's web services, first introduced in December 1993, continues to expand as more services are added. In FY 1998, 95% of searches against NCBI databases were conducted via the web. In addition to GenBank and PubMed, a host of new services and databases were added and new versions of others were released. NCBI staff continued to make access and usage easier with improved documentation and tutorials. Information about NCBI, its databases and services, data submission and update, and individual scientist’s research projects is readily available, as well as an ever-increasing number of search tools. The web server provides capabilities for Entrez and BLAST searches and submission by BankIt. Many other web servers have links to the NCBI server to conduct searches and obtain the latest GenBank records. At the end of FY 1998, NCBI’s site was averaging over 3,800,000 hits daily.

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 20 gigabytes of data, including daily GenBank and dbEST update files, are downloaded daily. There are 2,000 FTP requests per day, increasing to 5,000 at the time of the bimonthly GenBank release.

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 the servers and have an answer returned 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. Numerous improvements were made in two new releases of Sequin, NCBI's stand-alone submission tool. A wider variety of input files (PHRED/PHRAP/CONSED) are accepted that contain quality scores as well as sequences; the quality graphs are displayed in Sequin's graphical viewer. Also, genome records can be assembled from individual overlapping records already contained in GenBank. Complete documentation and a tutorial are maintained on the web site. The web submission tool, BankIt, now in its fourth year of use, is still heavily used, but with the growing use of Sequin, submissions entering via this route have dropped from about 65% to just over 50% in the final quarter of FY 1998.

NCBI uses a network of Unix-based symmetric multiprocessor servers to support its Internet database services. Each day NCBI generates over 60 gigabytes of data to send over the Internet in response to over four million user transactions (queries, similarity searches, and FTP downloads of data files). During FY 1998 many of these servers were upgraded to accommodate the increased demand and the increased size of the databases they serve. PubMed and Entrez are based on a three-tiered architecture: a first tier of web servers and a second tier of database servers provide the interactive services; a third tier of systems serve indexing, updating, maintenance, and development functions. Computing resources for all three tiers were increased and corresponding upgrades were made to the network infrastructure that supports PubMed and Entrez. In addition, NCBI's main router was upgraded and two new Ethernet switches put into service. In collaboration with OCCS, routers were configured to provide redundancy at both the router and switch levels to insure uninterrupted service.

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 at the NCBI are comprised of a multidisciplinary group of scientists who carry out research on fundamental biomedical questions at the molecular level by developing and utilizing mathematical, statistical and other computational methods. The approach is both theoretical and applied. These two lines of research are mutually reinforcing and complementary. The basic research has led to new practical methods and the application of these methods has opened new areas of research.

There have been a variety of applied and theoretical studies of biologically important molecules and their functions, as well as continued development and improvement of algorithms and statistics for their analysis. Computer program and algorithm development have included the development of protein sequence search methods using patterns as seeds; affine gap costs for sequence alignments; database design, data management and analysis for gene expression arrays; measures of threading specificity; mapping by electronic PCR; a system for the easy analysis of lots of sequences (SEALS); knowledge measures in multiple human relevance judgements; and Monte Carlo estimations of P-values in Markov chains. Macromolecular sequence analysis programs have been applied to investigate Wnt antagonists; phosphoesterase domains; HORMA domains in mitotic checkpoints; chromosome synapsis and DNA repair; toprim domains in topoisomerases, primases and nucleases; yeast histone H1; metalloenzymes, including phosphopentomutases, phosphoglycerate mutases, alkaline phospatases, and sulfatases; Gadd45 and MyD118 proteins; histone deacetylases; HIV-1 haplotypes and HTLV-I subtypes; vif genes in HIV-1; and, P-type ATPases.

Genome-scale projects have continued to be a staple of Computational Biology Branch research. Projects have included the analysis of POV1, a novel gene up-regulated in prostate cancer; finding systematic annotation errors in whole genome

annotations; the use of metabolic pathways for functional annotation of genome sequences; the analysis of prokaryotic genomes; genome analysis using clusters of orthologous groups of proteins sequences from whole proteomes; and patterns of evolutionary conservation in positionally cloned human disease genes.

Other efforts have included the development of a database of histones and histone fold sequences and structures; research on a model of the HIV infectivity; description of a measure for the success of protein fold recognition; and several studies on evolution, including the analysis of the translation machinery, non-coding sequence conservation in vertebrates, synonymous and non-synonymous substitution distances in mouse and rat genes, and a study of the extent and nature of contacts between protein molecules in crystal lattices.

The intramural group is engaged in over 40 projects, many of which involve collaborations with NIH and other research laboratories. The work is reviewed by a Board of Scientific Counselors of extramural scientists (see Appendix 4 for list of members). The high caliber of the work has been evidenced by the number of peer-reviewed publications, over 85 in FY 1998, and the requests for outside collaborations. The staff has made 78 presentations at major scientific meetings, academic departments, and companies engaged in molecular biology research; in addition, they made 22 presentations at NCBI's computational biology lecture series and hosted presentations by 45 invited speakers.

The Visitors Program continues to be successful in bringing members of the scientific community to the NCBI to engage in collaborative research with the Computational Biology Branch as well as joint activities in database design and implementation with the Information Engineering Branch. This program, administered in conjunction with Oak Ridge Associated Universities, facilitated 45 visits by 36 individual senior researchers this past year.

User Support and Outreach

As part of its mandate to facilitate the use of databases and software by the biology community, NCBI maintains a user support group staffed by scientists, librarians, and information specialists with broad experience in handling biology and medical information. The primary focus of this group is to support the particular services that NCBI offers by e-mail, phone, and fax. NCBI has extended its outreach to the library science community by invited presentations and workshops on biotechnology information topics.

As the number of database services and number of users has increased, the scope of user support services has also expanded. NCBI staff in the Information Resources Branch, with contractor support, provide responses to queries for information and assistance. The three main areas of user support include: information about GenBank and related molecular biology database services and data submission; technical assistance for submission of new GenBank data and revision of existing records; and technical assistance with Entrez and other data retrieval services. Most responses are immediate and nearly all answers or information are provided within 24 hours of receipt of a message. Authors who submit their sequences to GenBank are furnished with accession numbers for publication within 48 hours. During FY 1998, assistance with general PubMed inquiries from both users and publishers was provided by staff from MEDLARS Management Section, NLM.

To increase awareness of NCBI and its programs, NCBI staff participate in exhibits, seminars, workshops, and courses, both nationally and internationally. In FY 1998 NCBI staffed exhibits at scientific society meetings, including the American Society for Human Genetics; American Society for Cell Biology; Biophysical Society; American Society for Tropical Medicine and Hygiene; American Association for the Advancement of Science (AAAS); American Association for Cancer Research; The Protein Society; American Society for Microbiology; Special Libraries Association; American Crystallographic Association; American Society for Biochemistry and Molecular Biology; and, Association of American Medical Colleges. Additional exhibits were staffed at more specialized meetings, including TIGR/Computational Genomics, TIGR/Microbial Genomes, Genome Sequencing & Analysis Conference, Yeast Genetics and Molecular Biology, and NIH Research Day.

In addition, NCBI members participated as faculty at one or more courses sponsored by the American Association for Cancer Research (Aspen, Co., and Beijing); Catholic University of American; The Johns Hopkins School of Medicine; George Mason University's Institute for Computational Sciences and Informatics; Gordon Research Conference; The Jackson Laboratory; Marine Biological Laboratory; American Society for Biochemistry and Molecular Biology; Boston University/Museum of Science; Medical Library Association; Special Library Association; Massachusetts Health Science Library Network; and, the International Center for Genetic Engineering & Biotechnology in Trieste, Italy. NCBI also co-sponsored, with NHGRI, a mini-course in genomics that attracted an average of 75 NIH attendees at a series of lectures.

The NCBI newsletter was distributed to a mailing list of 25,000 biologists and institutions, and new fact sheets and tutorials on programs and services were distributed at all public forums where NCBI was represented.

The NCBI also participates in an advisory and collaborative role with other government agencies such as the Patent and Trademark Office and the Department of Agriculture on programs involving biotechnology information. Within the NIH, the NCBI coordinates with other institutes and particularly with the National Human Genome Research Institute and the National Cancer Institute on databases and informatics programs that impact information exchange on a national level.

Extramural Programs

Funding for extramural bioinformatics activities is the responsibility of another division of the NLM, the Extramural Program. It offers a program of grants for computer analysis of molecular biology data. A wide variety of work 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's informatics fellowship program.

Biotechnology Information in the Future

The explosive growth in the fields of genetics and molecular biology and the application of this knowledge to medical practice reinforce the need to build and maintain a strong infrastructure of information support. NCBI continues to be engaged in developing and employing new methods for disseminating knowledge to the biomedical community. Based on a core of advanced intramural research in several areas of computational biology, NCBI rapidly addresses the evolving informatics needs for genome research with state-of-the-art software and databases. Genomic information resources such as NCBI have repeatedly demonstrated their value as indispensable discovery tools for basic research. The value of these resources will only continue to grow as they support the breakthroughs in basic research and provide us with better understanding and treatment of human disease.

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 as grants-in-aid, and in some instances as contracts, to the extramural community in support of the goals of the National Library of Medicine. Review and award procedures conform to NIH policies.

The Grant Programs

EP issues grants in several categories for which the motif in general is Medical Informatics:

- Resources for Information Management
- Training
- Research and Research Resources
- Publication
- Educational Technology
- SBIR/STTR
- Other

Resource Grants

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 relative to their intent as well as dollar amounts and duration of time. They are considered "seeding" grants designed to initiate a resource or service or program, and to help such to eventually become self-sustaining. The commonalities among the four Resource Grants are that public and private, nonprofit health institutions/organizations engaged in education, research, patient care, and administration are eligible and that there must be library involvement.

Information Access Grants

Information Access Grants, aimed primarily at hospitals, clinics, community health centers and similar small health institutions and organizations, support computers to access Internet Grateful Med/PubMed and also fund the automation of public access catalogs for health science libraries. These grants provide up to $12,000 per institution participating and are available to single as well as multiple institutions working together.

Note: Awards preceded by an '*' are first year, or competing awards, all others are continuations of awards made in previous years.

*Pittman, Jan M., $84,457
2 G07 LM06280-02
Three Rivers AHEC, Columbus, GA

*Todini, Carole, $11,880
1 G07 LM06615-01
West KY AHEC Internet Access Training Project
Madisonville, KY

*Lang, Charles, $101,731
1 G07 LM06703-01
Project Inform—Information For Rural Medicine
Carl Vinson VA Medical Center, Dublin, GA

Information Systems Grants

Information Systems Grants, ranging up to $150,000 per year for up to three years, are aimed at larger hospitals and medical schools; they support computer access to information on a more complex level than Information Access Grants.

Seltzer, Adam, $137,116
5 G08 LM05478-05
Health Sciences Info. Network, Jackson, MS

*Dennis, Sharon E., $140,314
2 G08 LM05684-04
A Model Multimedia Support Center
University of Utah, Salt Lake City

**Denotes new award in FY 1998*

Eaton, Elizabeth, $92,014
5 G08 LM05774-02
Creating Digital Multimedia Resources
Tufts University , Boston

Jones, Michael O., $59,136
5 G08 LM06218-03
Computerized Archive of Am. Traditional Med.
University of California, Los Angeles

Mueller, Christine A., $153,597
5 G08 LM06239-02
Health Info. Access for Rural Nurse Practitioners
University Of Minnesota, Minneapolis

Matthew, Suzanne L., $172,695
5 G08 LM06492-02
Northwoods Healthnet
Northern Wisconsin AHEC, INC., Wausau

*Roderer, Nancy K., $128,986

1 G08 LM06575-01
New Haven Public Health Database
Yale University, New Haven

*Lemkau, Henry L., JR., $136,202
1 G08 LM06583-01
Point-Of-Care, Team-Based Information System
University Of Miami, Coral Gables

Internet Connections Grants

The Internet Connection Grants provide funding for single institutions up to $30,000 and multi-institutions up to $50,000 to initiate Internet access for health institutions by funding gateway/router equipment, Internet Service Provider fees, and line charges in the first year. Some institutions with existing Internet access extend it to other institutions.

*Coffee, James K., $40,440
1 G08 LM06403-01A1
Internet Connections for Medical Institutions
Greater Lawrence Family Health Center, MA

*McNeill, Kevin, $41,757
1 G08 LM06685-01
Internet for Rural Arizona Telemedicine Sites
University Of Arizona, Tucson

*Pretrick, Eliuel K., $29,564
1 G08 LM06690-01
Expanding Internet in Micronesia
Federated States/Micronesia Dept. of Human Svcs.
Palikir, Pohnpei FM

*Cho, Jung, H., $50,000
1 G08 LM06695-01
Camden County Internet Connection
Health & Human Services Department, Camden, NJ

*Burkett, James, $50,000
1 G08 LM06699-01
Internet Connection For McKennan Health Services
Mckennan Hospital, Sioux Falls, SD

*Lammers, Keith, $30,000
1 G08 LM06701-01
Internet Connection For Medical Institutions
Pennsylvania College Of Optometry, Philadelphia

*Malinowski, Barbara, $29,922
1 G08 LM06702-01
Internet Connection—Westfield Hospital
Westfield, NY

*Johnson, Kathryn E., $26,619
1 G08 LM06711-01
Healthcare Forum Internet Access Project
Healthcare Forum, San Francisco

*Smith, Melanie A., $9,074
1 G08 LM06713-01
Public Health Authority of Cabarrus County
Concord, NC

*Hengler, W. C., $29,783
1 G08 LM06714-01
Internet Connection for the Visiting Nurses Assn.
Butler, PA

*Astorino, Gina, $48,962
1 G08 LM06717-01
Internet For Region VIII Community Health Centers
Denver, CO

*McVeety, Katharine M., $50,000
1 G08 LM06738-01
Internet Linking of the MI Rural Health Services
Alpena General Hospital, Alpena, MI

*Berman, Audrey, $13,795
1 G08 LM06742-01
Bay Area Tumor Institute Internet Connection
Oakland, CA

*Schmidt, Kenneth, $49,895
1 G08 LM06745-01
Internet Access for Downeast Maine
Regional Medical Center at Lubec

*Forbes, Janet K., $11,045
1 G08 LM06748-01
Internet Access for Fayette County Health Dept.
Vandalia, IL

*Wilson, Stephen S., $16,935
1 G08 LM06760-10
Integration Of a Research LAN and the Internet
La Jolla Institute For Allergy, San Diego

*Brudno, Roger E., $28,850
1 G08 LM06762-01
Oroville Hospital Internet Connection Project
Oroville, CA

IAIMS Grants

Integrated Advanced Information Management Systems (IAIMS) Grants are designed to support institution-wide information systems that link a variety of individual and institutional databases and information systems for patient care, education, research, library, and administration. IAIMS Grants support two phases, planning and implementation, with the program goal being to support organizational mechanisms that foster the integration and sharing of

various information systems. The planning phase funds up to $150,000 for one to two years; the operational phase up to $500,000 per year for five years or $550,000 with an IAIMS apprenticeship option.

Phase I

*Panko, Walter B., $149,990
1 G08 LM06307-01A1
IAIMS Planning at UIC
University of Illinois, Chicago

*Russell, Floyd K., $146,364
1 G08 LM06277-01A1
Robert C. Byrd Health Sciences Ctr. IAIMS
West Virginia University, Morgantown

Phase II

Mitchell, Joyce A., $550,000
5 G08 LM05415-04
Missouri IAIMS
University Of Missouri, Columbia

Stead, William W., $550,000
5 G08 LM05443-05
Fast Track Provision of IAIMS
Vanderbilt University, Nashville

Paton, John A., $544,181
5 G08 LM05583-03
IAIMS Implementation at Yale
Yale University, New Haven

Fuller, Sherrilynne, $549,997
5 G08 LM05620-05
UWHSC Integrated Information Infrastructure
University Of Washington, Seattle

*Niland, Joyce C., $550,000
2 G08 LM005697-03
City of Hope National Medical Center
Duarte, CA

*Friedman, Charles, $550,000
1 GO8 LM06625-01
University of Pittsburgh

Training And Fellowship (MLAA)

Training in Informatics

Exploiting the potential of computers and telecommunication for health care information requires investigators who understand fundamental problems of knowledge representation, decision support, and human-computer interface. NLM remains the principal support nationally for research training in the fields of medical INFORMATICS, including clinical and basic science domains. NLM provides three mechanisms of support for its training activities.

Five-year institutional training grants support approximately 150 trainees at predoctoral and postdoctoral levels. Twelve institutions currently receive such support: NCI and NIDR contribute funds to NLM to help support slots at these training sites for applicants interested in radiation therapy and dental informatics respectively.

Shortliffe, Edward H., $659,060
5 T15 LM07033-15
Stanford University, CA

Gatewood, Lael, $306,279
5 T15 LM07041-15
University Of Minnesota, Minneapolis

Miller, Perry L., $394,510
5 T15 LM07056-12
Yale University School Of Medicine, New Haven

Friedman, Charles P., $564,276
5 T15 LM07059-12
University of Pittsburgh, PA

Downs, Stephen M. $491,966
5 T15 LM07071-07
University of North Carolina, Chapel Hill

Clayton, Paul D., $696,687
5 T15 LM07079-07
Columbia University, New York

Spackman, Kent A., $167,644
5 T15 LM07088-07
Oregon Health Sciences University, Portland

Mitchell, Joyce A., $448,156
5 T15 LM07089-07
University of Missouri, Columbia

Greenes, Robert, $1,147,348
5 T15 LM07092-07
Harvard Medical School, Boston

Gorry, G. A., $326,708
5 T15 LM07093-07
Training in Computational Biology and Medicine
William Marsh Rice University, Houston

Tierney, William M., $201,716
5 T15 LM07117-02
Regenstrief Medical Informatics Res. Fellowships
Indiana University, Bloomington

Gardner, Reed M., $408,230
5 T15 LM07124-02
University Of Utah, Salt Lake City

Individual informatics research fellowships are available to those who seek research training similar to offerings at the institutional training sites but a site of their choosing. Individual applied informatics fellowships are available to individuals who want to learn informatics techniques and technology for application in their current professional specialties.

Carter, Alexandre, $40,994
5 F31 LM00040-05
Minority Predoctoral Fellowship Program
Harvard Medical School, Boston

Soo Hoo, Kent, JR., $19,057
5 F37 LM00044-03
Content Based Indexing for Medical Image Mgmt.
University Of California, San Francisco, CA

Norris, Patrick R., $17,596
5 F37 LM00053-02
Therapeutic Goals in Intelligent ICU Monitoring
Vanderbilt University, Nashville

Nigrin, Daniel J., $68,770
5 F38 LM00055-02
Database Independent Analysis Of Endocrine Tests
Children's Hospital, Boston

Schoeffler, Katherine M., $30,996
5 F37 LM00062-02
Measurable Attributes for Controlled Terminologies
Duke University, Durham, NC

*Hau, Barbara, $47,572
1 F38 LM00063-01
Promoting Public Health: A Telecommunity Net.
Taos, NM

*Lehmann, Christopher, $45,500
1 F38 LM00064-01
Web-Based Case Simulation
Johns Hopkins University, Baltimore

*Padman, Rema, $69,100
1 F38 LM00065-01
Data Mining for Healthcare Decision Support
University of Pittsburgh, PA

*Schubart, Jane R., $67,224
1 F38 LM00068-01
Evaluation of an Info. Resource for Improving Care
University of Virginia, Charlottesville

*Bachman, Jean A., $65,000
1 F38 LM00069-01A1
Informatics Training to Impact Health of School
University of Missouri, St. Louis

*Terrazas, Enrique, $54,075
1 F38 LM00072-01
Web-Based Laboratory Order Entry System
University Of California, San Francisco

Education of Health Sciences Librarians in Informatics

After reviewing options in FY 1997, NLM elected to approach the issue with a three-element project beginning in FY 1998 of which two involved EP:

- All existing NLM Informatics Training Programs would be strongly encouraged to develop and offer training within the curriculum suitable for those interested in health science libraries. NLM agreed to provide additional funding for any slots awarded to librarians. Results have been very encouraging with librarians already in place or soon to begin at University of Pittsburgh, University of Missouri, and Yale University. Others are expected to become involved.
- Applied Informatics Fellowships were again widely publicized to those in the library community. No applications have been received.

Training of Minorities

NLM is participating in an NIH-wide fellowship program aimed at encouraging under-represented minorities in research careers. In FY 1998:

- NLM provided funding to the University of Wisconsin in the amount of $79,818 for the training of a minority applicant.
- An Internet Connection Grant was awarded to the University of Arizona to provide Internet access for several health centers including the Mariposa Community Health Center in Nogales which has a large Hispanic population.
- An Applied Informatics Fellowship was awarded to Barbara Hau of La Plaza Telecommunity in Tao, New Mexico, to create a "Virtual Community Health Resource Library" relevant to the health information needs of providers and multicultural (including Hispanic) consumers in Tao and Rio Arriba Counties in rural northern New Mexico.
- An Applied Informatics Fellowship was awarded to an Hispanic investigator, Enrique Terrazas, M.D., of the University of California, San

Francisco to develop a web-based laboratory order-entry system.

Research and Research Resources (PHS 301)

Research grants are made through a variety of mechanisms, including individual research projects, cooperative agreements, research resource grants and others.

Medical Informatics

NLM's research grants sponsor investigation of basic and applied medical-knowledge issues that arise at the intersection of biomedicine, computer science, and human behavioral sciences. NLM has been increasingly successful in recent years in interesting other NIH Institutions in supporting informatics projects wholly or in partnership with NLM.

To promote Informatics research projects of particular interest to nursing, a Program Announcement (PA-95B010), Enhancing Clinical Care Through Nursing Informatics, jointly supported by NLM and the National Institute of Nursing Research, was released December 2, 1994. Although NINR has subsequently refocused its funding priorities, the PA stands. Applications received because of it are reviewed by NLM's BLRC, and if meritorious, are funded by NLM and/or NINR.

Jaffe, Conrade, $235,021
5 R01 LM5007-09
Indexing Image Databases for Motion Similarity Retrieval
Yale University School of Medicine, New Haven

Berner, Eta S., $203,032
5 R01 LM05125-07
MD Use of Decision Support System Data
University of Alabama, Birmingham

*Jelliffe, Roger, $434,323
2 R01 LM05401-07
Decision Supports and Databases for Drug Dosage
University of Southern California, Los Angeles

Chute, Christopher G., $128,214
5 R01 LM05416-06
Latent Semantic Indexing to Support Data Retrieval
Mayo Foundation, Rochester, MN

Balas, E. A., $66,814
5 R01 LM05545-04
Meta Analysis of Clinical Information Service Trials
University of Missouri, Columbia

Selker, Harry P., $125,000
3 R01 LM05607-03S1
New Mathematical Models for Medical Events
New England Medical Center Hospital, Boston

Toga, Arthur W., $230,324
5 R01 LM05639-04
Digital Representation and Visualization of the Brain
University Of California, Los Angeles

*Musen, Mark, $537,276
2 R01 LM05708-04
Software Architecture: Guideline-Directed Therapy
STANFORD UNIVERSITY, CA

Lang, Walter, $142,827
5 R01 LM05917-03
MEDLINE and Computer Conferencing by Dentists
University Of Michigan, Ann Arbor

Marshall, Bryan, $224,691
5 R01 LM05997-02
Info. Integration and Virtual Therapy in the SICU
University of Pennsylvania, Philadelphia

Miller, Randolph A., $354,128
5 R01 LM06226-02
Patient Care Provider Order Entry with Support
Vanderbilt University, Nashville

Miller, Randolph A., $26,328
3 R01 LM06226-02S1
Patient Care Provider Order Entry with Support
Vanderbilt University, Nashville

Good, Walter F., $93,855
7 R01 LM06236-04
Non ROC Measures-Evaluating Image Compression
University Of Pittsburgh, PA

Bowman, Douglas M., $164,570
5 R01 LM06243-03
Spatial/Symbolic Brain Info, Management System
University of Washington, Seattle

Brennan, Patricia F., $319,670
5 R01 LM06249-02
Customized Computer Support for Home Care
University of Wisconsin, Madison

Brennan, Patricia F., $79,818
3 R01 LM06249-02
NIH Initiative for Minorities In Medical Informatics
University of Wisconsin, Madison

Huang, H. K., $280,977
5 R01 LM06270-02
Digital Hand Atlas in Assessment of Skeletal Dev.
University of California, San Francisco

Friedman, Carol, $204,002
5 R01 LM06274-02
Unlocking Data from Med. Records with Text Processing
City University, New York

Hersh, William R., $180,858
5 R01 LM06311-02
New Approach to Retrieval System Evaluation
Oregon Health Sciences University, Portland

Brinkley, James F., $296,107
5 R01 LM06316-02
Structure Based Visual Access to Biomedical Info.
University Of Washington, Seattle

*Sonnenberg, Frank A., $237,536
1 R01 LM06321-01A1
Decision Analytic Support for Clinical Guidelines
University of Med/Dent of New Jersey
Piscataway, NJ

Grobe, Susan J., $165,247
5 R01 LM06325-02
Automated Analysis of Intervention Narrative
University of Texas, Austin

Luther, Paul W., $143,575
5 R01 LM06326-02
Three-Dimensional Reconstruction of Synapses
University of Maryland at Baltimore

Goodwin, Linda, $247,266
5 R01 LM06488-02
Informatics & Perinatal Knowledge Building
Duke University, Durham

*Gustafson, David H., $419,029
1 R01 LM06533-01A1
Computer Information, Support: Breast Cancer
University Of Wisconsin, Madison

Haug, Peter J., $190,843
5 R01 LM06539-02
Semantic Parser For Medical Free Text
LDS Hospital, Salt Lake City

*Brodley, Carla E., $359,782
1 R01 LM06543-01A1
Content-Based Image Retrieval for Med. Databases
Purdue University, West Lafayette, IN

*Kohane, Isaac S., $299,716
1 R01 LM06587-01
Health Info. Identification and De-Identification Toolkit
Children's Hospital, Boston

*Gardner, Reed M., $516,201
1 R01 LM06591-01
Clinical Software Quality Review Process
LDS Hospital, Salt Lake City

*McKeown, Kathleen R., $302,301
1 R01 LM06593-01
Multimedia Summaries-Patient Status Post Bypass
Columbia University, New York

*Metaxas, Dimitri, $214,027
1 R01 LM06638-01
Analysis of #30 Cardiac Motion from Tagged MRI
University of Pennsylvania, Philadelphia

*Miller, Perry L., $272,102
1 R01 LM06682-01
Tools for Computer-Based Clinical Guidelines
Yale University School of Medicine, New Haven

*Cooper, Gregory F., $199,215
1 R01 LM06696-01
Modeling Cause and Effect from Clinical Data
University of Pittsburgh, PA

*Ezquerra, Norberto F., $343,723
1 R01 LM06726-01
Knowledge Discovery in Cardiac Imagebases
Georgia Tech Research Corp.
Atlanta, GA

*Neuwald, Andrew F., $360,556
1 R01 LM06747-01
Sequence-Based Prediction of Protein Function
Cold Spring Harbor Laboratory, NY

Shiffman, Richard, $111,212
5 R29 LM05552-02
Knowledge Processing for Clinical Guidelines
Yale University School of Medicine, New Haven

Lenert, Leslie, $112,700
5 R29 LM05626-05
Computer Interpretation of Free-Text Data
Stanford University, CA

Hripcsak, George, $117,687
5 R20 LM05627-05
Linking Knowledge-Based Systems to Clin. Databases
Columbia University, New York

Lehmann, Harold P., $125,131
5 R29 LM05647-05
Formalizing the Notion of Clinical Significance
Johns Hopkins University, Baltimore
Gorman, Paul N. $109,427

5 R29 LM05663-04
Assessment of Information Seeking in Primary Care
Oregon Health Sciences University, Portland

Yang, Yiming, $99,071
5 R29 LM05714-04
Mapping for Indexing and Retrieval of MEDLINE
Carnegie Mellon University, Pittsburgh

Johnson, Stephen B., $116,660
5 R29 LM05783-04
Medical Information through Natural Language
Columbia University Health Sciences, New York

Tong, David A., $83,300
5 R29 LM06004-03
Model Based Interpretation-Intracardiac Electrograms
University Of Texas Health Science Center
San Antonio, TX

Wagner, Michael M., $98,015
5 R29 LM06233-02
Belief Network-Based Reminder Systems That Learn
University of Pittsburgh, PA

Rutledge, Geoffrey W., $40,123
5 R29 LM06235-03
Advanced Computer Methods for ICU Care
Beth Israel Deaconess Medical Center, Boston

Langlotz, Curtis, $52,402
5 R29 LM06238-03
Computer-Based Explanation Methods for Decision Models
University of Pennsylvania, Philadelphia

Shahar, Yuval, $111,950
5 R29 LM06245-03
Knowledge-Based Temporal Abstraction of Clinical Data
Stanford University, CA

*Wong, Stephen T., $116,211
1 R29 LM06300-01A1
A Multimodality Neuroimaging Database System
University of California, San Francisco

Rolland, Jannick P., $93,646
5 R29 LM06322-02
3D Dynamic Anatomy—Virtual Reality Prototype
University of Central Florida, Orlando

Brummer, Marijn, $113,409
5 R29 LM06486-02
Interactive 4D Visualization of Heart Disease
Emory University, Atlanta
*Ohno-Machado, Lucila, $198,939
1 R29 LM06538-01
Tools for Connectionist Classification
Brigham and Women's Hospital, Boston

Sonnenberg, Frank A., $72,900
5 K04 LM00096-05
Knowledge Mgmt. for Clinical Decision Analysis
University of Med/Dent of New Jersey
Piscataway, NJ

Biotechnology Informatics (Computational Molecular Biology)

The techniques of informatics are indispensable tools for handling the complex data generated by molecular biology research. NLM continues to provide research grants for informatics projects in this area of basic medical science, as well as training grants, and grants for support of research resources. 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.

Because tools for sequence analysis are now well-developed and widely available, EP is expanding the computational biology grant program into related areas including molecular evolution, population studies, and other research areas opened up by the availability of sequence data.

Herzenberg, Leonore A., $318,146
5 R01 LM04836-08
FACS-Penguin-Knowledge Based Support for Flow Cytometry
Stanford University, CA

Myers, Eugene W., JR., $157,399
5 R01 LM04960-09
Software for the Analysis of Biosequences
University of Arizona, Tucson

Pearson, William R., $235,239
5 R01 LM04969-11
Comparison of Protein Sequences and Structures
University of Virginia, Charlottesville

Miller, Webb, $398,300
5 R01 LM05110-10
Software for Analyzing Biosequence Data
Pennsylvania State University, University Park

Brutlag, Douglas L., $222,416
5 R01 LM05716-05
Multiple Representations of Biological Sequences
Stanford University, CA

Roberts, Richard J., $155,597
5 P41 LM05800-04

Rebase - The Restriction Enzyme
New England Biolabs, Inc.
Beverly, MA

Altman, Russ B., $308,705
5 R01 LM06244-03
Representing Bio. Data for Molecular Modeling
Stanford University, CA
FY 1998 $6,208 Presidential Career Award

Ledley, Robert S., $1,175,092
5 P41 LM05798-04
Protein Information Resource
National Biomedical Research Foundation
Washington, DC

Markley, John L., $447,294
5 P41 05799-03
Biological Magnetic Resonance Data Bank
University of Wisconsin, Madison

Pagon, Roberta, $159,962
5 P41 LM06001-03
Helix-A Directory of Med. Genetics Laboratories
Children's Hospital and Medical Center, Seattle

Pagon, Roberta, $594,038
5 P41 LM06029-02
GENLINE-A Clinical Genetics Knowledge Base
University Of Washington, Seattle

Jurka, Jerzy W., $264,131
5 P41 LM06252-03
REPBASE-A Database of Repetitive Sequences
Genetic Information Research Institute
Los Altos, CA

Altman, Russ B., $111,937
5 R29 LM05652-05
Modeling/Computing with Uncertain Structures
Stanford University, CA

Goldstein, Richard A., $111,688
5 R29 LM05770-04
Computational Approaches to Protein Structure
University of Michigan, Ann Arbor

NLM also participates with 15 other organizations in the Human Brain Project, which 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.

Ellisman, Mark, Ph.D., $15,450
5 R01 DC03192-02
Development of a 3D Cell-Centered Neuronal Database
La Jolla, CA

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 monographs on special areas of biomedical research and practice; research monographs in the history of medicine and the life sciences; writings 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, but projects in electronic publishing, video, and other media were also supported. The program has an informal self-imposed ceiling on direct costs per grant per year. In November the ceiling was raised from $25,000 to $35,000.

Hast, Malcolm, $37,925
5 R01 LM05675-04
Annotated Translation of Vesalius-Fabrica
Northwestern University, Evanston, IL

Lunbeck, Elizabeth, $34,970
5 R01 LM05934-02
Practicing Psychoanalysis in the U.S., 1910-1915
Princeton University, Princeton, NJ

Bonner, Thomas N., $11,000
5 R01 LM06262-03
Biography of Abraham Flexner (1866-1959)
Wayne State University, Detroit, MI

*Schiebinger, Londa, $35,121
1 R01 LM06566-01
Fertility Control-17th & 18th Cent. Med. Botany
Pennsylvania State University, University Park

*Golden, Janet, $39,000
1 R01 LM06567-01
A History of Fetal Alcohol Syndrome
Rutgers University, Piscataway, NJ

*Kiple, Kenneth F., $38,500
1 R01 LM06574-01
History and Culture of Food and Nutrition
Bowling Green State University, OH

*Speaker, Susan, $35,285
1 R01 LM06590-01A1
Talking about Drugs…1875-1900
Middleburg, PA

*Pernick, Martin S., $36,929
1 R01 LM06662-01
What Is Death? Changing Meanings Since 1740
University Of Michigan, Ann Arbor

Educational Technology

No separate budget has been identified for this area, nor is NLM likely to fund simple development or demonstration projects. However, educational technology viewed from the viewpoint of educational research is of great interest. Applicants seeking funding for educational technologies are advised to structure their efforts as research projects for which R01 informatics grants would be appropriate.

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.

Frawley, Sandra, $49,732
3 R43 LM06330-01S1
Linking Web-Based Retrieval to Online Patient Record
Medical Decision Associates, Inc.
New Haven, CT

*Baclawski, Kenneth P., $90,298
1 R43 LM06665-01
Biomedical Science Info. Retrieval and Mgmt.
Softamp, Waltham, MA

*Liddy, Elizabeth D., $98,672
1 R43 LM06671-01
Medlink, Textwise LLC
Syracuse, NY

Liddy, Elizabeth D., $50,000
3 R43 LM06671-01S1
Medlink, Textwise LLC
Syracuse, NY

*Kalinyak, Judith E., $90,298
1 R43 LM06679-01
An Intranet System for Cardiovascular Disease
Intelligent Medical Objects
Northbrook, IL

*Doyle, Michael D., $100,000
1 R43 LM06728-01
Anatomical Mapping And Visualization System
Muritech, Inc.
Cambridge, MA

Other Grants

Conference Grants

Support for conference 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.

N0l LM 7-3529, $153,300
National Reference Center for Bioethics Literature
Georgetown University
Washington, DC

Other Extramural Programs Activities

HPCC and Outreach

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

In addition to its standing grant programs, Extramural Programs Division of NLM engages in a number of special projects aimed toward important biomedical goals, and often involving cooperation with another NIH institute or other Federal agency. Some examples of such activities in FY 1998 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 FY 1998 was $1,000,000, and represents an arm of the HPCC initiative. Total project budget from all sources may exceed $50 million over a 5 years. NLM is making available to interested applicants the Unified Medical Language System Knowledge Sources and the Visible Human datasets. Applicants are also free to use resources of their own choosing.

Program for Informatics Training in Africa

Computer-assisted information processing and communication have become critical to both medical research and health care in the developed world. Parts of the developing world are falling further and further behind technically during the explosive growth of these tools elsewhere, despite enormous opportunities they offer to improve science and health delivery in lower income nations.

During the development of NLM's long range international plan, NIH's Fogarty International Center (FIC) proposed that a training program in medical informatics be developed jointly by the Center and NLM. The broad aim of the program would be to support training that would improve the ability of developing country scientists and health professionals to use information technology to advance their work.

FIC and NLM held an international workshop at NIH on February 24, 1998 to help focus the specific objectives and structure of the program. As initial funding for the program is limited, FIC decided to focus initially on African scientists and health professionals. The reasons for this choice are several. Hurdles to development of informatics capability in Africa are significant, including connectivity, hardware, political and social barriers. However, the opportunities in Africa to leapfrog many of the traditional barriers to information access and communication are also great. EP participated in the development of the program, and conducted peer review of the applications received.

Improving Public Health and Health Services Research

Because of the remarkable potential of information technology to process huge quantities of data, there is growing interest by both professionals and the public in measuring the quality and effectiveness of current medical practices. Noting that the Agency for Health Care Policy and Research (AHCPR) recently funded a number of training sites intended to increase expertise in health services research, NLM initiated a series of joint discussions and planning for a workshop to explore ways and means of using informatics, and in particular, the NLM informatics research training programs, to facilitate the AHCPR training goals, and to further our mutual interest in health services research and public health research. A jointly sponsored symposium on the topic will be held in FY 1999 at NLM.

Informatics for the National Heart Attack Alert Program

The NHAAP was established by the National Heart, Lung, and Blood Institute in June 1991 to promote the use in clinical practice of scientific research that indicated that prompt medical treatment—notably, thrombolytic therapy—significantly improves survival rates for patients with acute myocardial infarction (AMI) and improves the quality of life for the patients and those around them. However, progress has been slow, and at present only a small minority of patients capable of benefiting are actually receiving the medications. The NHAAP has identified three stages where delay can occur in the identification and treatment of individuals with a potential heart attack:

Stage I: Patient and bystander recognition of the symptoms and signs of AMI and their actions in response to these symptoms.

State II: Prehospital action by emergency medical services providers - that is, the response to patients prior to their arrival at the hospital.

Stage III: Hospital action by health care providers at the hospital to identify and treat patients with the symptoms and signs of AMI.

Following a symposium sponsored jointly by NHLBI, NLM, and the Agency for Health Care Policy and Research, NLM published a Request for Proposals intended to obtain contract research and development services related to the use of medical informatics as an approach to reducing or eliminating some or all of the various obstacles which are hindering the ability of the NHAAP to reach its goals. NHLBI transferred $800,000 to NLM to support this program. Eight contracts for Phase One

planning contracts were awarded. It is expected that subsequent phases will include modeling and implementation contracts.

Lucille Ohno-Machado, M.D., Ph.D.
Decision Systems Group
Boston, MA

G. Octo Barnett, M.D.
Massachusetts General Hospital
Boston, MA

Henry C. Chueh, M.D.
Massachusetts General Hospital
Boston, MA

Victor J. Strecher, Ph.D., MPH
University of Michigan
Ann Arbor, MI

Stavroula Osganian, M.D., Ph.D.
New England Research Institutes, Inc.
Watertown, MA

Harry P. Selker, M.D., MSPH
New England Medical Center
Boston, MA

Pat McLaughlin, M.D.
University of Missouri-Columbia
Columbia, MO

James Cimino, M.D.
Columbia University
New York, NY

Miscellaneous Special Projects

NLM continues to transfer funds to other agencies to support projects of broad scope and utility for biomedical research.

Y1-LM-8035-01, $150,000
Coordinate Funding and Oversight of a Database to Contain Atomic Coordinates and Related Structural Information for Biological Macromolecules
Brookhaven National Laboratory
Uptown, NY

Y1-LM-8040-01, $50,000
CarbBank: A Structural and Bibliographic Database for Complex Carbohydrates
Department of Energy
University of Georgia

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 issued 156 awards for FY 1998. We are now able to e-mail the Notice of Grant Awards to the Grantee Institution instead of mailing hard copies. We are also now using the new Information Management Planning Analysis and Coordination II (IMPAC II) system to issue our Notice of Grant Awards. Effective October 1, 1998, there will be an increase in the stipend levels for the National Research Service Awards.

Review Committee Activities

NLM's scientific merit peer review group, the Biomedical Library Review Committee (BLRC) met four times in 1998 and reviewed 135 applications of which 80 were approved. The Committee (see Appendix 5 for roster of members) operates as a "flexible" review group; i.e., it is composed of three standing subcommittees: nine members on the Medical Library Resource Subcommittee, seven members on the Medical Informatics Subcommittee; and four members on the Biomedical Information Subcommittee. The subcommittees consider research applications in medical library projects, medical informatics, and biotechnology information respectively.

A final peer review of applications is performed by the Board of Regents, which meets three times a year, approximately three months after the Biomedical Library Review Committee. (A roster of the Board 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 of "specials" 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.

Review Reform

The NIH is participating in the President's Reinventing Government initiative. A portion of this activity has been the establishment of a Peer Review Oversight Group charged with coordinating, evaluating, and making policy recommendations for all peer review conducted at NIH. The Group is continuing to meet and will consider input from the applicant community. Some changes which have occurred include:

- The Division of Research Grants has become the Center for Scientific Review, and a number of the standing study sections have been reorganized.
- Five specific criteria have been chosen as a basis for assessing the merit of NIH research projects grants.
- CFIRST (R29) awards for newly independent biomedical investigators will become R01 (traditional Research Project) awards after June 1998. Applicants wishing to apply for a FIRST award will indicate this on the cover of the application.
- New guidelines will go into effect after October 1, 1998 for the inclusion of children in NIH funded research involving human subjects.

Divisional Operations

An increase in the operating budget was requested for the Internet Connection Program. EP is moving toward the electronic processing of grant review for both the Board of Regents and Biomedical Library Review Committee meetings.

Administrative and Personnel

EP relocated to the Rockledge One Building in June 1998. EP continues to incorporate the new NIH streamlining initiative by utilizing appropriate staff in various tasks throughout the year. EP filled two Grants Technical Assistants and one program analyst positions.

Summary

NLM's EP, like almost all extramural grant divisions at NIH, regrets that not all applications of good quality can be funded, but the grants which can be made are furthering NLM goals in most key areas. However, support for developing the educational technology of Informatics remains uncomfortably small, and, most importantly, we have not yet expanded the informatics research budget commensurate with the increase in informatics scientists leaving our training programs.

Table 9

Extramural Grant and Contract Program
(dollars in thousands)

Category	*FY 1996*		*FY 1997*		*FY 1998*	
	No.	*$*	*No.*	*$*	*No.*	*$*
Resource projects	**53**	**5,754**	**40**	**5,307**	**39**	**5897**
IAIMS	(10)	(3,011)	(8)	(3,040)	(9)	(3,690)
Access	(5)	(476)	(6)	(320)	(3)	(198)
Systems	(12)	(1,387)	(11)	(1,412)	(10)	(1,352)
Connections	(26)	(880)	(15)	(535)	(17)	(657)
Research	**68**	**14,817**	**80**	**14,994**	**72**	**15,887**
Medical informatics projects	(24)	(4,433)	(32)	(5,955)	(38)	(8,384)
Medical informatics resource	(1)	(1,419)	(1)	(230)	…	…..
Biotechnology	(19)	(5,264)	(27)	(6,259)	(16)	(4,730)
Cooperative agreements	(7)	(1,872)	(2)	(754)	…	…..
Career awards	(16)	(1,638)	(18)	(1,796)	(17)	(1,773)
Library science	(1)	(191)	…	…..	…	…..
Digital library	…	…..	…	…..	(1)	(1,000)
Training	**19**	**4,501**	**20**	**5,645**	**23**	**5,519**
Institutional	(10)	(4,178)	(12)	(5,290)	(12)	(4,993)
Fellowship	(9)	(323)	(8)	(355)	(11)	(526)
Publications	**12**	**327**	**9**	**247**	**8**	**269**
Bioethics	**1**	**458**	**1**	**498**	**1**	**513**
SBIR/STTR	**2**	**196**	**4**	**401**	**5**	**508**
Regional Medical Library	**8**	**6,283**	**8**	**6,611**	**8**	**6,710**
NIH Tap		**849**		**830**		**830**
Totals:	**163**	**$33,185**	**162**	**$34,533**	**156**	**$36,133**

Office of Computer and Communications Systems

Roy A. Standing
Acting Director

Application Support Services

One of the major services performed by OCCS is to provide operational systems for applications that support the mission of the NLM. This includes requirements definition, analysis, system development, and operational support. Over the past 25 years numerous custom-built systems have been developed and implemented. These legacy systems continue to provide most of the support for traditional library services, and unique NLM services and information retrieval of many of NLM's databases.

OCCS ensures these applications are operational during all scheduled hours (typically around the clock) and responds to change requests from the program areas. These legacy systems support acquisition, cataloging, circulation, preservation, binding, online public access, and document delivery of the world's historical and contemporary biomedical literature; development of a thesaurus and classification system used to organize bibliographic information; authoritative indexing and cataloging of material; building and dissemination of bibliographic databases.

Ensuring these legacy systems are fully operational while at the same time resolving Y2K issues and supporting NLM's System Reinvention efforts has been especially challenging.

Application Development

MEDLINE*plus*

MEDLINE*plus* is an easy-to-understand resource for the public which includes MEDLINE as well as links to NIH consumer health organizations, clearinghouses, health-related organizations, self-help groups, and clinical trials. MEDLINE*plus* is the result of a joint effort by OCCS and LO. In the initial development of MEDLINE*plus*, pages were developed by LO using static HTML pages. It soon became clear that this approach would be extremely difficult to create and maintain. OCCS was asked to come up with an alternative.

The result was a system utilizing a commercial database (Oracle) and ColdFusion for the development of the web pages. A key feature of the design is that the web pages are dynamically built from the database. This allows the content providers to concentrate on content rather than being concerned about HTML formatting. The database design and a working prototype was developed within a few weeks in order to be prepared for a scheduled meeting with public librarians.

The MEDLINE*plus* prototype was presented to representatives from 37 libraries on July 27th. Design and development of the project has continued in order to provide an input and approval system that can support remote contract work. This has included creating a static page generation application that parallels the dynamic input system. The application was moved to a production environment with a new production ColdFusion server accessing a production Oracle database server. Link checking is performed to ensure all links are valid.

World Wide Web Support Services

OCCS is responsible for the development and support of the NLM home page(s) that provides links to other NLM services and numerous other sites. This includes the computing and networking environments. To provide a stable environment for this dynamic site OCCS designed and implemented a dual development/production environment. NLM web contributors are able to develop and update HTML pages in the development environment with no impact on production. Upon approval, appropriate pages are promoted to the development systems. OCCS developed scripts to support this promotion process.

The NLM home page is constantly being evaluated to ensure it is the best it can be. Early in 1998 the new NLM Intranet "look and feel" design was implemented. This included restructuring to support the use of Frames, Stylesheets, Tables, and image maps. Use of the Internet and NLM Intranet continues to grow dictating changes in computing hardware and networking facilities to support this workload and to provide a more robust environment.

Loansome Doc

A Web-based server to support registration and submitting interlibrary loan requests from PubMed and Internet Grateful Med was put in production. This effort required coordination from many NLM components and from the Regional Medical Libraries.

NLM System Reinvention

Integrated Library System

A team of OCCS and LO staff conducted a multi-year analysis of commercially available Integrated Library Systems(ILS). This analysis was completed in February 1998, when NLM selected

Voyager, developed by Endeavor Information Systems Inc., as the Library's new integrated library system to support its basic library functions. Voyager, an integrated information management system designed for academic and research libraries, will be used by NLM for acquisitions, serials control, cataloging, collection management, circulation, preservation, binding, and an Online Public Access Catalog (OPAC). The OPAC will provide the retrieval engine for online access to the Library's cataloging records for monographs, audiovisuals, and serials, replacing existing online access mechanisms such as Locator, CATLINE, AVLINE and SERLINE. Voyager is a fully integrated system that combines open system architecture and relational database technology. It will allow NLM to incorporate data from the ILS into its other unique applications. Voyager will replace numerous internal custom-built systems developed at NLM over the past 25 years.

Following the award OCCS and LO staff worked with Endeavor to add new features to the product, including enhancements to serials processing and enhancements to the closed stack request module and a binding module. The process of extracting information from NLM's disparate legacy systems into a single integrated system was extremely challenging requiring the combined efforts of over 100 NLM staff members. Hundreds of computer programs were developed to support this effort.

The first phase of the ILS implementation will begin later this year with the release of Voyager for in-house use by NLM staff for cataloging and acquisitions work.The Web-based Voyager OPAC will be available for public use in early 1999. As noted above the OPAC will provide the retrieval engine for online access to the Library's cataloging records for monographs, audiovisuals and serials. The implementation of Voyager was a key milestone in NLM's System Reinvention efforts. Numerous other applications, including DOCLINE, SERHOLD and the creation of journal article citations (e.g., MEDLINE) will rely on the bibliographic data in the Voyager ILS.

DOCLINE Reinvention

Successful completion of this project will result in a replacement system for DOCLINE, NLM's Interlibrary Loan System. The goal is to re-engineer the current DOCLINE system utilizing state-of-the-art technology while adding functionality. Key features will be a Web access with interfaces to PubMed and the Voyager Integrated Library System.

The project includes DOCLINE itself and two key other components: SERHOLD and DOCUSER. SERHOLD is the system where participating DOCLINE libraries submit their holding information to be used by the DOCLINE system when routing interlibrary loan requests. DOCUSER is the system where information about participating DOCLINE libraries is maintained. The system is being developed using ColdFusion Web development tool and Oracle. Development is expected to be completed in the summer of 1999 with full implementation in the fall of 1999.

Data Creation and Maintenance of MEDLARS Databases

A major component of NLM's Systems Reinvention effort is the creation of a computing system to support the creation of MEDLINE and other databases. The system must provide journal authority control, handle both print and electronic material, support for the addition of MeSH terms to each article, and a mechanism to maintain the MeSH terms consistent with the current year of MeSH.

The Voyager Integrated Library System and the MeSH2000 thesaurus maintenance system are essential components of the new data creation and maintenance system. Process flow and data analysis have been completed and the resulting data model was used to create a database (Oracle). The database was populated with MEDLINE records and prototypes have been created. Work continues with a projected completion date of fall 1999.

Thesaurus Development

Central to the creation of all MEDLARS database are the Medical Subject Headings. A project (MeSH2000) to create a new thesaurus maintenance system has been under way for some time. Originally conceived to simply re-engineer the legacy system, the scope of the project was expanded to permit MeSH to be more in line with the UMLS Metathesaurus. This change in scope resulted in restructuring of the database design as well as rethinking of the process model.
Java, Oracle and Oracle Context are the key development tools being utilized. Development was completed late in 1998; system and acceptance testing will be completed in the spring of 1999.

RELAIS

The final v1.5 release of Relais was installed and tested in April. Requests for Overnight Photocopy Service(OPS) was the first group of requests to be processed. In the middle of May, we began downloading all NLM DOCLINE requests into Relais for processing. In June, we began downloading all NIH library electronic delivery

requests from DOCLINE and Loansome Doc into Relais.

CUSTOMERQ and HELPQ

CUSTQ is a commercially available software product that provides centralized support for reporting tracking and problem resolution. This system went operational and supports the centralized customer support services for much of NLM.

HELPQ is a commercially available software product to support OCCS online trouble reporting was installed and implemented. It went into production in June 1998

Local Area Network and Communications

During FY 1998, work continued on the project of upgrading the NLM LAN communications from 10BaseT Ethernet to 100BaseT Ethernet using a switched environment. The core of the LAN was first upgraded to this capacity, and then conversion of user subnets was started.

The conversion and migration to the GroupWise e-mail and calendar system from cc:mail and Quickmail was completed in FY 1998. This has simplified and improved e-mail administration by reducing the number of systems supported to two—Unix (POP, Eudora, IMAP) and GroupWise.

The NAL (Novell Application Launcher) was deployed for use as a software distribution tool via the LAN. NAL was used to roll out the GroupWise client software. Eventually NAL will totally replace Saber/McAfee for network based menus. The NAL was used for the automated distribution of Voyager software, McAfee Virus protection, and other applications via the LAN to staff PCs.

Preliminary investigations were performed for Y2K compliance for PCs, servers, and network devices and software. Some remediation was completed, and monitoring of vendor solutions for COTS products was started.

The IT functions of the Extramural Programs Office were supported in EP's relocation to the Rockledge I Building in Bethesda. On-site technical support was provided for the PC, network, and IMPAC II systems.

Overall NLM Network Management was improved with the addition and use of management and monitoring tools. One such tool is the Netscout RMON probe that is attached to the T3 link to our Internet Service Provider B GTE/BBN Planet. This enables us to monitor the traffic and utilization of the traffic on this important link.

Y2K

OCCS is meeting the Y2K compliance mandates set forth by the Office of Management and Budget and the Department of Health and Human Services. NLM had earlier identified 14 IT systems as having potential Y2K problems. Of these 14, 3 were considered critical to NLM's mission and required to undergo Independent Validation & Verification (IV&V) testing.

OCCS is in the process of replacing the library automation functions by a Y2K compliant COTS system—Voyager Integrated library System. The MEDLARS information retrieval system has been made Y2K compliant and will be replaced by PubMed in 1999. PubMed is Y2K compliant. The remaining mainframe systems (database creation and maintenance, DOCLINE, MESH and publications) have been made Y2K compliant by OCCS program development efforts. The TOXNET system has been replaced with a Y2K compliant system**.** The IV & V for these systems is scheduled to be completed by March 31, 1999.

NLM Computer Center

In addition to the normal operating system and program product support functions, the major accomplishments of the OCCS Mainframe Systems Team were related to the various efforts required for Year 2000 (Y2K) compliance. These Y2K efforts consisted of three phases. Phrase 1 was a successful hardware upgrade (disk drives and memory) that was required in order to provide concurrent support for both IGM and Y2K testing. Phase 2 tasks included the installation of the following software on a dedicated Y2K virtual computer: OS/390 Operating System, numerous IBM programs products, numerous non IBM products, and the MEDLARS Shell and Supporting Tasks. Phase 3 consisted of the development and successful execution of Y2K test plans for all of the phase 2 software components in the dedicated Y2K environment. In summary, work accomplished during 1998 left the OCCS mainframe and software supported by the Systems Team well positioned for Y2K.

Computer Operations

The in-house cartridge tape library was updated and consolidated, allowing for the removal of over 5,000 outdated and damaged cartridges.

Several program entities have transferred their systems into the NLM computer facility, including the Specialized Information Systems which transferred TOXNET system. Those organizations have also been provided with floor space, electrical power (UPS), system monitoring, system interaction and system reporting on a 24 hour, 7 days a week

basis, allowing OCCS to move towards total facilities management.

OCCS was successful in getting CBOMS (a mainframe job scheduler) to be Y2K compliant and compatible with the current OS390 mainframe operating system.

Information Systems Laboratory

The Information Systems Laboratory (ISL) has as its primary responsibility the maintenance and support of OCCS's 40+ UNIX servers and workstations, as well as the NT workstations in the NLM public reading room. In addition, during 1998, the ISL was responsible for the following major developmental tasks:

- Establish an Oracle infrastructure to support production and developmental databases, and serve as the Oracle Database Administrator (DBA) for these databases.
- Improve overall NLM computer security. This was done by forming the NLM Computer Security Committee to: provide a forum for sharing computer security information; act as an NLM Computer Incident Response Team; and develop NLM wide computer security policies. Also, to improve security, ISL installed both network intrusion detection software to monitor Internet traffic for hacker activity, and network scanning software to probe NLM systems for security weaknesses.
- Define the hardware environment and perform the initial system configuration for NLM's new Voyager Integrated Library System.
- Create a new High Availability/High Performance infrastructure to support NLM's WWW services.
- Configure and maintain NLM's Images of the History of Medicine WWW server.

User Services

- The Desktop Computer Support contract was awarded to Digicon Corp. of Bethesda. The contract commenced on June 15, 1998. They are performing the work previously done by Trawick and BDM, whose contracts each ended on June 19, 1998.
- As part of the OCCS HelpDesk reinvention, the OCCS HelpDesk was renamed to Information Technology Services Center (IT Services Center).
- Digicon Corporation, in conjunction with MIL Corp., played a substantial role in the Y2K audit of NLM's workstations and servers. Over 1,200 units were tested in less than two months.
- NT training courses were offered to the NLM by OCCS.
- 16 Dell Notebooks were ordered for the mobile training OCCS is providing. Classes are taught by Laszlo Nagy of MIL Corp and consist of MS Word, Excel, Powerpoint and Access. Future courses may include Netscape, HTML and HelpQ.
- 269 new 400 MHz workstations were configured and installed at NLM.

Administration

Executive Officer
Donald C. Poppke

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 FY 1998:

Internet Grateful Med (IGM): Several databases, including TOXLINE and BIOETHICSLINE, were added to IGM towards the end of the fiscal year. At the same time, IGM began using the new PubMed system in place of the old ELHILL retrieval system for all its MEDLINE searches.

Document Delivery Systems: NLM's Loansome Doc service, through which users can request copies of journal articles from a local medical library, was upgraded this year. The new web-based Loansome Doc is linked to both PubMed and IGM, giving users of either system a simple way of immediately ordering a copy of an article found in a search. Another system, Relais, installation of which began in FY 1997, went into full operation in May 1998. Relais streamlines NLM's document delivery process and is particularly efficient at electronic transmission of copies of articles requested by NLM's users.

Integrated Library System (ILS): NLM purchased Endeavor Information System's Voyager ILS product and contracted with the company to make certain enhancements to the product to meet NLM's needs. Besides working out the details of the changes to be made to the product, members of NLM's project team got extensive training in the operation and use of Voyager. By the end of the fiscal year, NLM staff had also completed the bulk of a major effort to clean up and reorganize the catalog and other existing library system files in preparation for transferring them to the new system.

Financial Resources

In FY 1998, the Library had a total appropriation of $160,516,000. Table 11 displays the FY 1998 budget authority plus reimbursements from other agencies, and the allocation of these resources by program activity.

TABLE 10

Financial Resources and Allocations, FY 1998 (dollars in thousands)

Budget Authority:	
Appropriation, NLM	$160,885
Plus: Reimbursements	10,037
TOTAL	**$170,922**
Budget Allocation:	
Extramural Programs	$ 36,033
Intramural Programs	126,214
Library Operations	(64,812)
Lister Hill National Center for Biomedical Communications	(36,877)
National Center for Biotechnology Information	(15,221)
Toxicology Information	(9,304)
Research Management and Support	8,675
TOTAL	**$170,992**

The 1998 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 1997, Carol Hotton, Ph.D., received a Postdoctoral Intramural Research Training Award from the National Center for Biotechnology Information (NCBI). Dr. Hotton received her Ph.D. in botany from the University of California. Immediately before her appointment with NCBI, Dr. Hotton served as a Research Associate at the National Museum of Natural History in Washington, D.C. While at NCBI, Dr. Hotton's research includes the continuing analysis of the taxonomy of species in GenBank, the advancing of systematics of angiosperms, caryophylles, and organismal evolution at the molecular level.

In November 1997, Sharee Pepper, Ph.D., was selected for the position of Health Scientist Administrator within the Division of Extramural Programs. Formerly, Dr. Pepper worked for NIDDK as a Health Scientist Administrator for more than six years. In her new position, Dr. Pepper is responsible for planning, conducting, and coordinating the scientific and technical merit review of grant applications assigned to NLM.

In November 1997, Yuri Wolf, Ph.D., was appointed as an NCBI Visiting Fellow. Dr. Wolf received his Ph.D. in bioinformatics at the Institute of Cytology and Genetics, Russian Academy of Sciences, Novosibirsk, Russia. At NCBI, Dr. Wolf

will conduct research in the analysis of intergenome relationships.

In November 1997, Zuoming Deng, Ph.D., received a Postdoctoral Intramural Research Training Award from the NCBI. Dr. Deng received his Ph.D. from the University of Texas at Houston. Dr. Deng is involved in the analysis of DNA and protein sequences.

In January 1998, the NLM entered into an Intergovernmental Personnel Act agreement (IPA) with Neil Rambo, Associate Director, Pacific Northwest Regional Medical Library, University of Washington, Seattle. Mr. Rambo has significant experience in health information services, outreach, and training in the use of advanced information services. Mr. Rambo is developing public health informatics training materials, organizing pilot tests of the materials with groups of public health professionals identified by Regional Medical Libraries, and modifying and updating the materials as indicated by the pilot tests.

In January 1998, Paul A. Fontelo, M.D., M.P.H., joined NLM as a Special Expert with the Office of High Performance Computing & Communications. Dr. Fontelo received his M.D. from the University of the Philippines. He is a board certified pathologist and has served as Chief, Telepathology Division, Armed Forces Institute of Pathology; Chief, Department of Pathology, 196th Station Hospital SHAPE, Belgium; and Chief, Department of Early Diagnosis, U.S. Army Medical Research, Research Institute of Infectious Diseases. Dr. Fontelo has had extensive experience in computer system hardware and software integration and high-speed networking. As a Special Expert with NLM, Dr. Fontelo is working on developing NLM's Next Generation Internet in health and medical areas.

In February 1998, Thomas L. Madden, Ph.D. was selected for a Research Biologist position with the Information Engineering Branch, NCBI. Dr. Madden received his Ph.D. in physics from the University of California, Santa Cruz and did postdoctoral work at Brandeis University. Dr. Madden started at NCBI as a postdoctoral fellow in 1993. Dr. Madden maintains and manages the BLAST database. He has extensively redesigned the popular sequence database search tool which is used by scientists all over the world thousands of times a day and has led to innumerable scientific discoveries.

In February 1998, Junga Kim was appointed as a Visiting Associate with the Information Engineering Branch, NCBI. Ms. Kim received her Master's Degree in biology from George Washington University in 1993 and came to NCBI as a Scientific Data Analyst immediately afterwards. She serves as the assistant GenBank coordinator, providing day-to-day detailed supervision of the sequence indexers. Ms. Kim also works on troublesome sequence submissions and updates and has created web pages for in-house use to help maintain the scheduling of various quality assurance duties performed by the NCBI staff.

In March 1998, NLM's Division of Specialized Information Services (SIS) announced the selection of Jeanne C. Goshorn and Michael D. Moore for two branch chief positions. Ms. Goshorn will direct the Biomedical Information Services Branch and Mr. Moore will direct the biomedical Files Implementation Branch. Ms. Goshorn joined the NLM in 1980 as a Technical Information Specialist. Before becoming branch chief, Ms. Goshorn was responsible for managing a variety of activities and services related to the TOXLINE family of databases, the RTECS database and query response, training, and exhibit efforts of the Toxicology Information Program. Mr. Moore joins the NLM from the Department of Commerce, Patent and Trademark Office where he served as Chief of the Scientific and Technical Information Center, Biotechnology/Chemical Division, directing the Center's daily operations and long-term information systems and services, and supervising the development of online biotechnology databases.

In May 1998, Fernando Burbano, Director Office of Computer and Communication Systems (OCCS) left the NLM for a position with the State Department. Roy Standing, Chief, Information Management Branch, OCCS is serving as Acting Director, OCCS.

In June 1998, Paul A. Kitts, Ph.D., joined NLM as a Special Expert with the Information Engineering Branch, NCBI. Originally from Great Britain, Dr. Kitts received his Ph.D. in genetics from the University of Glasgow, Scotland. Dr. Kitts came to NLM from CLONTECH Laboratories Inc. in Palo Alto, California, where he spent six years as a research scientist. There he developed new commercial reagents such as cloning vectors, expression vectors, reporter vectors and supporting resources. As a Special Expert with NCBI, Dr. Kitts will use his scientific expertise in the development of vectors and his experience in the biotechnology industry to provide scientific oversight of the design of GenBank and the content of the vector database.

Between June and September 1998, 10 appointments to Staff Scientist positions were made in NCBI under the newly expanded appointment mechanism for the appointment of scientists at NIH under 42 U.S.C. Sections 209 (g) and 209 (h) and C.F.R. Part 61B. This appointment authority applies to scientific positions in both the NIH intramural and extramural programs.

In July 1998, the OCCS announced the selection of Karen B. Casey and Wei Ma for two section chief positions. Ms. Casey, who serves as Chief, Information Collection Section, joined NLM

from the Information Resources Division, U.S. Coast Guard, where she was responsible for various applications systems, most notably the Marine Safety Information System and the Merchant Mariner Licensing and Documentation System. Ms. Ma, who serves as Chief, Software Development Branch, joined NLM from AT&T Corporation where she served as Senior Technical Specialist and Project Leader for a variety of projects including the AT&T web-based Customer Direct Platform. Ms. Ma brings to NLM expertise in software and systems analysis, design, integration, testing, and deployment.

In August 1998, Milton Corn, M.D., was appointed to the Senior Executive Service position of Associate Director for Extramural Programs. For the past eight years, Dr. Corn has served as Acting Associate Director, EP. During this time, he has contributed significantly to NLM's extramural activities. In his new appointment, Dr. Corn will continue to provide technical advice to the Director, NLM, on issues related to the Library's extramural program.

In August 1998, Karen Hajarian was appointed to the position of Director, MEDLARS Promotion, in the Office of the Chief, Bibliographic Services Division, Division of Library Operations (LO). Ms. Hajarian has an extensive background in nursing and over 20 years of national and international sales and marketing experience in the health care and information science industry. Ms. Hajarian had previously served as a Special Expert in LO promoting biomedical online information systems and providing solutions for Internet and World Wide Web access to electronic information. Before her association with the NLM, she held the position of Director, BRS Search Services with BRS Information Technologies, where she directed outside sales.

In August 1998, Diane Boehr was appointed as Cataloging Unit Head, Cataloging Section, Technical Services Division, LO. Ms. Boehr received her MLS from the University of Maryland, and for the past several years she has been an adjunct professor there. For the past 15 years, Ms. Boehr has served as a library services consultant for Costabile Associates, Inc. As Cataloging Unit Head, Ms. Boehr is responsible for planning, organizing, and providing administrative supervision for the unit.

In September 1998, Donald C. Poppke was appointed to the Senior Executive Service position of NLM Associate Director for Administrative Management. He had been with the NLM since 1995 as the Library's Executive Officer. Mr. Poppke received his Masters of Science in Technology Management from the American University in 1983. He previously worked as a Program Analyst for the National Cancer Institute, as Executive Officer for the National Center for Nursing Research, and as Chief, Public Health Branch in the Division of Public Health and Social Services Budget Analysis under the Office of the Secretary, HHS.

In September 1998, Terry S. Yoo, Ph.D., joined the staff of NLM as a Computer Scientist with the Lister Hill National Center for Biomedical Communications (LHNCBC), Office High Performance Computing and Communications. Dr. Yoo received his B.A. degree from Harvard University and his Ph.D. in Computer Science from the University of North Carolina at Chapel Hill in 1996. Prior to his appointment with NLM, Dr. Yoo served as Assistant Professor in the Department of Radiology at the University of Mississippi Medical Center in Jackson, MS. As a senior Computer Scientist with NLM, Dr. Yoo is responsible for conducting research projects to develop prototype systems and comprehensive models in medical informatics and digital imaging systems.

In September, Ms. Jennifer Marill was appointed as a Senior Systems Librarian in the Office of the Chief, Public Services Division, LO. Ms. Marill received a Masters in Russian and East European Studies from the University of Michigan Center for Russian Studies and a Masters in Library and Information Science from the University of Illinois Graduate Library School. For the past 6 years, Ms. Marill served as a Systems Librarian for Technical Services and Collection Development at the Washington Research Library Consortium, and prior to that, she served as a Senior Automation Planning Analyst with the Library of Congress.

Retirements

In November 1997, Thelma G. Charen, Technical Information Specialist, Division of Library Operations, NLM, retired after completing 53 years of Federal service, most of which she served at NLM. Ms. Charen has been an integral part of the NLM, having helped formulate the principles under which the Library provides subject access to the world's biomedical literature. Over the decades, she has trained hundreds of librarians and indexers in the use of Medical Subject Headings, the Library's controlled vocabulary thesaurus, and in the principles of subject control and access.

In January 1998, Donald R. Buckner, Ed.D., Materials Development Officer, LHNCBC, retired after serving over 30 years in the Federal sector. Dr. Buckner's career included serving as an adviser and policy development expert to identify research and development opportunities to solve communications and information transfer problems in the health sciences. Dr. Buckner also provided strong leadership in the Equal Employment Opportunity arena, both at the NLM and within the NIH community.

In January, 1998, Gerard T. Guthrie, Computer Systems Analyst, Biomedical Files

Implementation Branch, SIS, retired with 36 years of service. Mr. Guthrie joined NLM in 1981 and served as a senior computer systems analyst with primary responsibility for the requirements analysis that translate program needs into functional specifications for supporting SIS files on the ELHILL system, including CHEMLINE, TOXLINE and ChemID.

In May 1998, Mr. Robert Kicklighter, Chief, Database Administration Section, Information Management Branch, OCCS, retired with 30 years of service from the Federal government. Mr. Kicklighter began his Federal career in 1968 and joined NLM as a computer programmer in 1974. As Chief, Database Administration Section he was responsible for the management of operational software systems and the administration of all operational databases.

Awards

The NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to two employees: Dr. Stephen Bryant (NCBI) in recognition of outstanding work in designing the Molecular Modeling Database and for developing innovative approaches for protein structural comparisons; and Dr. George Thoma (LHNCBC) for his outstanding leadership in designing and implementing an automated data entry system for producing MEDLINE citation records.

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 award was Ms. Karen Sinkule, Division of Library Operations, for her foresight in the planning and implementation of a conservation and book repair lab, thus enabling scholars to use more items in the collection today and for many years to come.

The NIH Director's Award was presented to the following two individuals and team: Ms. Duane Arenales for superior management of the NLM's collection of biomedical literature; Ms. Patricia Carson in recognition of her resourcefulness and superb leadership in organizing major national and international events; and Dr. Mark Boguski and Dr. Gregory Schuler as members of the NCI Tumor Gene Index Project Team, in recognition of their scientific leadership in guiding the development of the Tumor Gene Index.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to three employees: Ms. Lois Ann Colaianni for exceptional leadership and innovative contributions to the Library's programs and services during her highly successful tenure as NLM's Associate Director for Library Operations; Mr. James Ostell for creative design and development of essential biomedical information resources; and Ms. Patricia Williams, for outstanding and innovative administrative support for NLM's basic library services.

The NIH Merit Award was presented to seven employees: Mary Smith, Office of Administration; Gale Dutcher, SIS; Dr. George Thoma, LHNCBC; and Kathi Canese, Rhonda Allard, Esther Baldinger, and Ione Auston, LO. Ms. Smith was recognized for exceptional management and effective administration of the contracting function at NLM. Ms. Dutcher was recognized for exceptional leadership and creativity in establishing programs to optimize awareness and availability of the NLM HIV/AIDS information services. Dr. Thoma was recognized for his leadership of advanced digital imaging systems projects and document management systems. Ms. Canese was recognized for contributions to the successful implementation of PubMed as NLM's primary MEDLINE retrieval system. Ms. Allard was recognized for her role in the development of new customer service tools needed to support free web access to MEDLINE. Ms. Baldinger was recognized for her leadership and technical expertise in coordinating and controlling the flow of journal information needed to support MEDLINE and other NLM online files. Ms. Auston was recognized for enhancing NLM's bibliographic services in health services research.

The PHS Commissioned Corps Commendation Medal was awarded to Dr. Richard Rogers for advanced demonstration of network-based information contributions in the promotion of research and health care application and Internet based teleconferencing technology, and to Dr. Michael Weisberg for contributions in the management of the activities of The Learning Center for Interactive Technology.

The NIH Harvey J. Bullock, Jr. Award for Equal Opportunity Achievement was presented to Mr. Michael Bumbray for his outstanding supervisory efforts in affording staff all possible opportunities to grow professionally and to work in a discriminatory-free workplace.

The NIH Quality of Work Life Award was presented to the four NLM employees: Mr. Donald Poppke for important contributions to the Library's Quality of Work Life Initiatives and strong commitment to ensuring the success of these activities; Ms. Sheila Levy for taking the initiative to improve staff morale, employee satisfaction and group cohesiveness by reorganizing space for employee offices necessitated by loss of space; and Ms. Alice Jacobs and Ms. Christa Hoffman as a team for their effort in providing support for the ergonomic evaluation of the Technical Services Division work stations and staff work habits and their application of the results in the redesign of the Cataloging Section's workstations.

TABLE 11

Staff, FY 1998 Full-Time Equivalents

Program	*Full-Time Permanent*	*Other*
Office of the Director	13	5
Office of Public Information	6	2
Office of Administration	50	5
Office of Computer and Communications Systems	54	9
Extramural Programs	14	3
Lister Hill National Center for Biomedical Communications	67	7
National Center for Biotechnology Information	31	14
Specialized Information Services	25	2
Library Operations	234	36
TOTAL	**494**	**83**
TOTAL FTEs	**577**	

NLM Diversity Council

Cassandra Allen, Chair
Public Services Division

In January, the NLM Diversity Council installed its first new members: LaShaun Alexander, Michael Bumbray, Perlita Liwanag, Joseph Pagano, and Frances Truong joined the Council for terms of two years beginning in January 1998. The vacancies they filled were created when four original members (Redmond Barnes, Liem Nguyen, Rose Marie Woodsmall, and Theodore Youwer) ended their terms. The fifth vacancy resulted from the retirement of Brenda Swanson in February 1997. The Diversity Council lost Mr. Pagano in February 1998 and Ms. Alexander in August 1998 when they accepted positions outside of NLM. Julia Royall joined the Council in August 1998. Current members of the Council are Cassandra Allen, Evelyn Bain, Michael Bumbray, Lou Knecht, Perlita Liwanag, Alexander Nobleman, Julia Royall, Kristine Scannell, Frances Truong and Monique Young. The Council's ex-officio members are David Nash, Donald Poppke, and Nadgy Roey.

Council accomplishments for this year include:

- Updating the Diversity Council web page to include a list of dates of religious and ethnic holidays that may warrant scheduling consideration, a link to the names of the members of NIH's Diversity Council, a link to information about NIH's Workplace Diversity Initiative, and the names of NLM Diversity Council members. The page is regularly updated to include information about Council events and projects.
- Conducting a diversity assessment by surveying NLM staff. The survey characterized the current NLM workforce, identified issues of concern to the workforce, and enabled the Council to make recommendations to address these concerns. The survey will also serve as a baseline which the Council will use to measure progress and track the make up of the NLM workforce and to identify changing concerns.
- Sponsoring a monthly brown bag session for all NLM staff and employees on diversity issues and issues of general concern. This activity also supports NLM's Quality of Work Life Strategy. The sessions offered in FY 1998 included: NIH Retirement Seminar; an encore showing of the NLM Diversity Training Video from August 1997 (this will be an annual showing targeted at new NLM staff and those wishing to refresh their skills); and Success Strategies for the 21st Century.
- Coordinating the viewing of two career-oriented videos for students from Calvin Coolidge High School as part of NLM's adopt-a-school program. The videos that the students saw were titled, "Me! A Librarian" which featured librarianship as a career, and "Connected: Career for the Future" which focused on different international careers.
- Recommending the establishment of the NLM Director's Employee Education Fund to assist employees in achieving educational goals. The fund will support up to two academic courses annually for NLM employees who wish to pursue their education.

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.nnlm.nih.gov/mar

2. **SOUTHEASTERN/ATLANTIC REGION**
University of Maryland at Baltimore
Health Science and Human Services Library
601 Lombard Street
Baltimore, MD 21201
(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.nnlm.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.nnlm.nih.gov/gmr

4. **MIDCONTINENTAL REGION**
University of Nebraska Medical Center
Leon S. McGoogan Library of Medicine
600 South 42nd Street
Omaha, NE 68198-6706
(402) 559-4326 FAX (402) 559-5482
States served: CO, KS, MO, NE, UT, WY
URL: http://www.nnlm.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.nnlm.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.nnlm.nih.gov/pnr

7. **PACIFIC SOUTHWEST REGION**
University of California, Los Angeles
Louise M. Darling Biomedical Library
Box 951798
Los Angeles, CA 90024-1798
(310) 825-1200 FAX (310) 825-5389
States served: AZ, CA, HI, NV and U.S. Territories in the Pacific Basin
URL: http://www.nnlm.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.nnlm.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:

ALBRIGHT, Tenley E., M.D. (Chair)
Two Commonwealth Avenue
Boston, MA

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

BOND, Enriqueta, Ph.D.
President
Burroughs Wellcome Fund
Durham, NC

FONSECA, Raymond J., D.M.D.
Department of Oral and Maxillofacial Surgery
Univ. of Pennsylvania School of Dental Medicine
Philadelphia, PA

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

FULLER, Sherrilynne, Ph.D.
Acting Director, Informatics
University of Washington School of Medicine
Seattle, WA

GAGE, John
Director, Science Office
Sun Microsystems Computer Corporation
Palo Alto, CA

KLEIN, Michele, MSLS
Systems Director, Library Services
Children's Hospital of Michigan
Detroit, MI

LEDERBERG, Joshua, Ph.D.
President Emeritus
Rockefeller University
New York, NY

PARDES, Herbert, M.D.
Vice President for Health Sciences
Dean of Faculty of Medicine
College of Physicians and Surgeons
Columbia University
New York, NY 10032

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 Sciences

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:

BUCHANAN, Bruce G., Ph.D. (Chair)
Professor of Computer Science
University of Pittsburgh
Pittsburgh, PA

CLAYTON, Paul D., M.D.
Professor of Medical Informatics
University of Pittsburgh
Pittsburgh, PA

JAFFE, Conrade C., M.D.
Director, Center for Advanced Instructional Media
Yale University School of Medicine
New Haven, CT

KAHN, Michael G., M.D., Ph.D.
Assistant Professor of Medicine
Division of Medical Informatics
Washington University
St. Louis, MO

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

WILKERSON, LuAnn, Ed.D.
Director, Center for Educational Development
UCLA School of Medicine
Los Angeles, CA

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:

ROBERTS, Richard J., Ph.D. (Chair)
Research Director
New England Biolabs
Beverly, MA

BUETOW, Kenneth H., Ph.D.
Chief, Laboratory of Population Genetics
National Cancer Institute
Bethesda, MD

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

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

MATISSE, Tara Cox, Ph.D.
Assistant Professor
Laboratory of Statistical Genetics
The Rockefeller University
New York, NY

PACE, Norman R., Ph.D.
Distinguished Professor of Biology
Indiana University
Bloomington, IN

SCHLICK, Tamar, Ph.D.
Associate Professor
Chemistry Department
New York University
New York, NY

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:

KULIKOWSKI, Casimir A., Ph.D. (Chair)
Professor of Computer Science
Rutgers University
New Brunswick, NJ

ALLMAN, Robert M., M.D.
Professor of Radiology
Univ. of Maryland School of Medicine
Baltimore, MD

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

BASLER, Thomas G., Ph.D.
Professor and Director
Libraries and Learning Resource Centers
Medical University of South Carolina
Charleston, SC

BROADNAX, Lavonda
Automation Operations Coordinator
Library of Congress
Washington, DC

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

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

FLORANCE, Valerie, Ph.D.
Director, Edward G. Miner Library
Asst. Professor, School of Medicine and Dentistry
University of Rochester
Rochester, NY

FRIEDMAN, Richard B., M.D.
Medical Director
Waianae Comprehensive Health Center
Waianae, HI

FUCHS, Rainer T., Ph.D.
Director, Bioinformatics
Glaxo Wellcome Research and Development
Research Triangle Park, NC

HOLST, Ruth
Director, Library Services
Columbia Hospital
Milwaukee, WI

HUANG, H.K., DSC
Professor and Vice Chairman
Director, Radiological Informatics
University of California at San Francisco
San Francisco, CA

LATTMAN, Eaton E., Ph.D.
Professor and Chair
Department of Biophysics
Johns Hopkins University
Baltimore, MD

MOLHOLT, Pat A., Ph.D.
Assistant Vice President
Columbia University Health Sciences
New York, NY

MOULT, John, Ph.D.
Professor
Center for Advanced Research in Biotechnology
Rockville, MD

NILAND, Joyce C., Ph.D.
Director of Biostatistics
City of Hope National Medical Center
Duarte, CA

ORTHNER, Helmuth , Ph.D.
Professor, Department of Medical Informatics
University of Utah Health Sciences Center
Salt Lake City, UT

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

RINDFLEISCH, Thomas
Director, Lane Medical Library
Stanford University
Stanford, CA

TANG, Paul C., M.D.
Medical Director, Information Systems
Northwestern Memorial Hospital
Chicago, IL

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:

ROLETT, Ellis L., M.D. (Chair)
Professor of Medicine
Dartmouth-Hitchcock Medical Center
Lebanon, NH

CABELLO, Felipe C., M.D.
Dept. of Microbiology and Immunology
New York Medical College
New York, NY

CLEVER, Linda Hawes, M.D.
Chair, Dept. of Occupational Health
California Pacific Medical Center
San Francisco, CA 94120

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

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

EPSTEIN, Neal, M.D.
Cardiology Branch
National Heart, Lung, and Blood Institute
Bethesda, MD

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

MAKINEN, Ruth H.
Head, Technical Services
University of Minnesota
Bio-Medical Library
Minneapolis, MN

MATHIEU, Alix, M.D.
Professor of Anesthesia
University of Cincinnati
College of Medicine
Cincinnati, OH

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

STRICKLAND, Ora Lea, Ph.D.
Professor, School of Nursing
Emory University
Atlanta, GA

WILLIAMS, Benjamin T., M.D.
Professor of Information Science
University of Illinois
Champaign, IL

INDEX

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 home pages of four of NLM's popular Web sites (May 2000).

NATIONAL INSTITUTES OF HEALTH

NATIONAL LIBRARY OF MEDICINE

Programs and Services
Fiscal Year 1998

U.S. Department of Health and Human Services

Public Health Service

Bethesda, Maryland

Contents

Appendixes

Tables

Preface

We at the National Library of Medicine seem to be operating on "fast forward" these days. Fiscal Year 1998 saw the introduction of new services and dramatic increases in the usage of existing services. More than 100 million MEDLINE searches were done this year compared to 8 million last year. The steeply rising trend in Web usage of MEDLINE via the Internet Grateful Med and PubMed shows every indication of continuing unabated. In April, we convened a group of 15 health sciences librarians to identify problems and suggest modifications with IGM and PubMed. The systems are constantly being improved.

Also this year there was an emphasis on how we can better serve the general public. For the first time, we added selected consumer health newsletters to MEDLINE. We met in July with representatives of 35 public library systems to plan how the NLM can help them meet the health information needs of their patrons. Early in FY 1999 a pilot program to do this will be announced along with a new NLM Web-based service to link the public with reliable health information. NLM also co-sponsored a "train the trainer" program to see if we could make MEDLINE and other electronic health information resources more widely accessible to seniors.

A new feature introduced by the Library this year is *Profiles in Science*. This Web service brings together the best in archival practices with state-of-the-art technology to present to the public a look behind the scenes of scientific findings and the unpublished writings, letters, and lab notes of great scientists. The first scientist selected was Oswald Avery, whose research laid the groundwork for modern genetics and molecular biology. Early in FY 1999 we will add the papers of Nobel Laureate Joshua Lederberg.

A notable advance in document delivery took place this year with the installation of "Relais," a modified commercial system that automates many interlibrary loan functions formerly done manually. Networked scanners and special workstations allow the NLM to turn around requests for copies of articles within a day, often within hours of a request being submitted. Some 3 million requests a year are entered by more than 3,000 libraries into DOCLINE, NLM's automated routing system. Almost 375,000 ended up at NLM and Relais is helping to handle this load.

My thanks to the staff and to our library colleagues around the world are tinged with sadness this year. Lois Ann Colaianni, Associate Director for Library Operations for 14 years, has announced her intention to retire at the end of 1998. She is widely know and respected throughout the community of health science librarians, and we will miss her greatly.

_____

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