

**Medical Library Association and Association of Academic Health Sciences Libraries  
Statement on FY21 Appropriations for the National Library of Medicine**

**Submitted to the Senate Subcommittee on Labor, Health and Human Services, and  
Education, and Related Agencies**

I, Mary M. Langman, Director, Information Issues and Policy, **Medical Library Association** (MLA), submit this statement on behalf of MLA and the **Association of Academic Health Sciences Libraries** (AAHSL). MLA is a global, nonprofit, educational organization with a membership of more than 400 institutions and 3,000 professionals in the health information field. AAHSL supports academic health sciences libraries and directors in advancing the patient care, research, education and community service missions of academic health centers through visionary executive leadership and expertise in health information, scholarly communication, and knowledge management.

We thank the Subcommittee for the opportunity to submit testimony supporting appropriations for the National Library of Medicine (NLM), an agency of the National Institutes of Health (NIH), and recommend \$479.7 million for NLM in FY21, a 5% (\$22.7 million) increase.

Working in partnership with the NIH and other Federal agencies, NLM is the key link in the chain that translates biomedical research into practice, making the data and other results of research readily available to all who need it. NLM is taking on additional responsibilities for NIH-wide efforts in data science and open science. As health sciences librarians who use NLM's programs and services every day, we can attest that NLM resources literally save lives. Therefore, investing in NLM is an investment in good health. NLM addresses Congressional priorities through rapid deployment of resources related to health emergencies including response to the COVID-19 pandemic and the opioid crisis, by providing clinical trial information, genomic sequencing data, and public access to research literature.

**Leveraging NIH Investments in Biomedical Research**

NLM's budget supports information services, research, and programs that sustain the nation's biomedical research enterprise. In FY2020 and beyond, NLM's budget must continue to be augmented to support modernization and expansion of its information resources, services, research, and programs which collect, organize, and develop new ways to make readily accessible rapidly expanding biomedical knowledge resources and data. NLM maximizes the return on investment in research conducted by the NIH and other organizations. It makes the results of biomedical information accessible to researchers, clinicians, business innovators, students, and the public, enabling such data and information to be used more efficiently and effectively to drive innovation and improve health. Rapid growth of data also necessitates funding that will ensure long-term sustainability of these valuable information resources.

NLM plays a critical role in NIH's data science and open science initiatives. NLM leads the development, maintenance and dissemination of key standards for health data interchange that are now required of certified electronic health records (EHRs). NLM builds, sustains, and augments a suite of almost 300 databases which provide information access to health

professionals, researchers, educators, and the public. It supports the acquisition, organization, preservation, and dissemination of the world's biomedical literature. In FY 2019, NLM made genomic sequence data available in the cloud. NLM's Sequence Read Archive (SRA) is the world's largest publicly available repository of next-generation genome sequence data, with more than 9 million records comprising 25 petabytes of data. To improve access and utility of SRA data, NLM uploaded the public access SRA data to two commercial clouds that have agreements with NIH's Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative. This transition significantly expands the discovery potential of the data. Freed from the limitations of local storage and computational resources, users are empowered to compute across the full corpus of SRA data without having to download and store large volumes of data. Moving to cloud platforms also makes it possible to develop customized tools and methods for asking research questions of the data.

### **Growing Demand for NLM's Information Services**

Each day, more than 6 million people use NLM websites and download 115 terabytes of data. Thousands of researchers and businesses submit 15 terabytes of data daily. Annually, NLM information systems process more than six billion human requests and eight billion computer-to-computer interactions. NLM's information services help researchers advance scientific discovery and accelerate its translation into new therapies; provide health practitioners with information that improves medical care and lowers its costs; and give the public access to resources and tools that promote wellness and disease prevention. Every day, medical librarians across the nation use NLM's services to assist clinicians, students, researchers, and the public in accessing information to save lives and improve health. Without NLM, our nation's medical libraries would be unable to provide quality information services that our nation's health professionals, educators, researchers and patients increasingly need.

NLM's data repositories and online integrated services such as GenBank, dbGaP, Genetics Home Reference (GHR), PubMed, and PubMed Central (PMC) are revolutionizing medicine and ushering in an era of personalized medicine. GenBank is the definitive source of gene sequence information. Each month, 2.1 million users accessed consumer-level information about genetics from GHR, which contains more than 2,700 summaries of genetic conditions, genes, gene families, and chromosomes. PubMed, with more than 30 million references to the biomedical literature, is the world's most heavily used source of bibliographic information with almost 1.35 million new citations added in FY19 and approximately 2.5 million users each day. NLM also launched a new PubMed platform for an improved user experience, including a new search algorithm with relevance rankings and better tools for citations. PubMed Central is NLM's digital archive which provides public access to the full-text versions of more than 6.1 million biomedical journal articles, including those produced by NIH-funded researchers. On a typical weekday more than 2.5 million users download more than 2.8 million articles.

NLM continually expands biomedical information services to accommodate a growing volume of relevant data and information and enhances these services to support research and discovery. NLM ensures the availability of this information for future generations, making books, journals, technical reports, manuscripts, microfilms, photographs and images accessible to all Americans,

irrespective of geography or ability to pay, and guaranteeing that citizens can make the best, most informed decisions about their healthcare.

### **Improving Public Access to Federally Funded Research Results**

The Department of Health and Human Services (DHHS) continues to work with NLM to ensure free public access to the results of taxpayer-funded research. HHS operating divisions, and ten other Federal agencies, use NLM's PubMed Central (PMC) as a common repository to provide access to peer-reviewed publications resulting from their research. In FY 2019, NLM added 600,000 full-text articles to PMC and continued linking articles to associated data by aggregating data citations, data availability statements, and supplementary materials. Since featuring these data links more prominently, daily downloads of supplementary material have increased by 30 percent. A subset of about 3 million articles in PMC is available for bulk retrieval for text mining and other research purposes.

### **Disseminating Clinical Trial Information**

*ClinicalTrials.gov*, the world's largest clinical trials registry, now includes more than 320,000 registered studies and summary results for more than 39,000 trials. As health sciences librarians who fulfill requests for information from clinicians, scientists, and patients, we applaud NIH and NLM for implementing requirements for clinical trials registration and results submission consistent with the FDA Amendments Act of 2007, and for applying them to all NIH-supported clinical trials. These efforts increase transparency of clinical trial results and provide patients and clinicians with information to guide health care decisions. They also ensure biomedical researchers have access to results that can inform future protocols and discoveries.

### **Partnerships Ensuring Outreach and Engagement in Communities Across the Nation**

NLM's outreach programs are essential to the MLA and AAHSL membership and to the profession. Through the National Network of Libraries of Medicine (NNLM), with over 7,000 members nationwide as of FY 2019, NLM educates medical librarians, health professionals, and the general public about its services and provides training in their effective use. The NNLM serves the public by promoting educational outreach for public libraries, secondary schools, senior centers and other consumer settings, and its outreach to underserved populations helps reduce health disparities. NLM's "Partners in Information Access" provides local public health officials with online information that protects public health.

Since May 2018, the NNLM has partnered with the NIH *All of Us* Research Program to support community engagement efforts by United States public libraries and to raise awareness about the program. To date, 376 libraries and 101 community-based organizations across 33 states and the District of Columbia have held 1,135 events with more than 40,000 people. Via the NNLM *All of Us* Community Engagement Network, 654 libraries support health literacy, including offering health and wellness programming in their communities. To support public library staff in providing health programming to the community, NNLM has also provided 491 classes, training, and workshops that have served 11,188 library staff.

NLM's MedlinePlus provides consumers with trusted, reliable health information on 1,000 topics in English and Spanish. It attracts more than 1 million visitors daily. NLM continues to enhance MedlinePlus and disseminate authoritative information via the website, a web service, and social media. MedlinePlus and MedlinePlus en Español have been optimized for easier use on mobile phones and tablets. *NIH MedlinePlus Magazine* and *NIH MedlinePlus Salud* are available in doctors' offices nationwide, and NLM's MedlinePlus Connect enables clinical care organizations to link from their EHR systems to relevant patient education materials.

### **Strengthening Data Science and Open Science Capacity**

NLM is a leader in data science and open science, including the acquisition and analysis of data for discovery and the training of biomedical data scientists. The library aims to strengthen its position as a center of excellence for health data analytics and discovery, and to spearhead the application of advanced data science tools to biological, clinical and health data. NLM is building a workforce for data-driven research and health by funding PhD-level research training in biomedical informatics and data science. The library also partners with NIH to ensure inclusion of data science and open science core skills in all NIH training programs, and is expanding training for librarians, information science professionals, and other research facilitators. NLM is participating in NIH-wide efforts to foster a culture that advances science and ensures the development and retention of a diverse, safe, and respectful workforce for data-driven research and health well into the future.

### **Responding to the Novel Coronavirus (COVID-19)**

The health sciences library community thanks Congress for providing NLM with the \$10 million supplemental appropriations to prevent, prepare for, and respond to the Coronavirus. NLM has been responding to COVID-19's rapidly evolving situation through its suite of tools and deep well of expertise in managing large and complex datasets and making them accessible to the public. Our frontline healthcare providers use NLM's databases to access the latest research datasets, literature publications, and scientific information about COVID-19. For example, NLM is:

- Making immediately available to the public in PubMed Central tens of thousands of coronavirus-related research publication and data contributed by major publishers
- Contributing to the COVID-19 Open Research Dataset (CORD-19), which represents the most extensive machine-readable coronavirus literature collection available for text mining to date, with more than 30,000 full-text scholarly articles from PMC as of mid-May 2020. The Text REtrieval Conference (TREC)-COVID Challenge makes use of the CORD-19 dataset to help search engine developers evaluate and optimize their systems in meeting the needs of the research and healthcare communities.
- Providing the biomedical community free and easy access to genome sequences from the coronavirus through the GenBank sequence database.
- Providing information about US clinical trials related to COVID-19 via ClinicalTrials.gov, which is also now making available information about trials listed in the World Health Organization's international clinical trial registry.

- Extending standard terminologies to include terms related to COVID-19, including codes for laboratory tests, chemical entities, and indexing terms.
- Applying machine learning techniques to research conducted at NLM to assist in identifying COVID-19 in X-rays and to identify and categorize relevant published literature.

### **Supporting Biomedical Informatics Research and Health Information Technology Innovation**

NLM conducts and supports informatics research, training and the application of advanced computing and informatics to biomedical research and healthcare delivery. NLM's National Center for Biotechnology Information (NCBI) focuses on genomics and biological data banks, and the Lister Hill National Center for Biomedical Communications (LHC), is a leader in clinical information analytics and standards. Many of today's biomedical informatics leaders are graduates of NLM-funded informatics research programs at universities nationwide. A number of the country's exemplary electronic and personal health record systems benefit from findings developed with NLM grant support. A leader in supporting the development, maintenance, and free, nationwide dissemination of standard clinical terminologies, NLM partners with the Office of the National Coordinator for Health Information Technology to support the interoperability of EHRs. NLM also develops tools to make it easier for EHR developers and users to implement accepted health data standards and link to relevant patient education materials. In FY 2019, NLM played a critical role in the development, usage, and utility of a data exchange standard to improve flow and availability of data, the Health Level Seven International (HL7) Fast Healthcare Interoperability Resources (FHIR®). NIH is encouraging funded investigators to use the FHIR standard to capture, integrate, and exchange clinical data for research purposes and to enhance capabilities to share research data. NIH has also announced to the small business communities its special interest in supporting applications that use FHIR in the development of health IT products and services. To support these efforts, NLM is managing the development and testing of FHIR tools that researchers can use to increase the availability of high-quality, standardized research datasets and phenotypic information for genomic research and genomic medicine.

We look forward to continuing this dialogue and thank you for your efforts to support funding of at least \$479.7 million for NLM in FY21, with additional increases in future years.