

Rising Expectations: Access to Biomedical Information

D. A. B. Lindberg, B. L. Humphreys

National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

Summary

Objective: To provide an overview of the expansion in public access to electronic biomedical information over the past two decades, with an emphasis on developments to which the U.S. National Library of Medicine contributed.

Methods: Review of the increasingly broad spectrum of web-accessible genomic data, biomedical literature, consumer health information, clinical trials data, and images.

Results: The amount of publicly available electronic biomedical information has increased dramatically over the past twenty years. Rising expectations regarding access to biomedical information were stimulated by the spread of the Internet, the World Wide Web, advanced searching and linking techniques. These informatics advances simplified and improved access to electronic information and reduced costs, which enabled inter-organizational collaborations to build and maintain large international information resources and also aided outreach and education efforts. The demonstrated benefits of free access to electronic biomedical information encouraged the development of public policies that further increase the amount of information available.

Conclusions: Continuing rapid growth of publicly accessible electronic biomedical information presents tremendous opportunities and challenges, including the need to ensure uninterrupted access during disasters or emergencies and to manage digital resources so they remain available for future generations.

Keywords

Access to information, genetic databases, digital libraries, consumer health information, clinical trials

Geissbuhler A, Kulikowski C, editors. *IMIA Yearbook of Medical Informatics 2008. Methods Inf Med* 2008; 47 Suppl 1: 165-72

Introduction

Today's world has high expectations regarding access to biomedical information. People expect up-to-date, easy, fast, reliable, and affordable access to a wide spectrum of information from basic scientific data to published literature to health information written for the general public. Of course, perceptions of what constitutes up-to-date, easy, fast, reliable, and affordable change over time and, within any given time period, may vary by geographic location as well as by individual. In the 21st century, "reliable" has the connotation of permanently accessible and uninterrupted by disasters or emergencies. To an increasing number of people, "affordable" means free - at least to them.

Despite persistent unevenness, access to biomedical information improved steadily over the past 50 years and dramatically over the past two decades. Improvements in access have been fueled by developments in computing and telecommunications, in informatics research, in inter-organizational cooperation, and in public policy. 2007 was a year of key anniversaries and significant new events in a story of continuing progress. Here are some of the milestones and challenges that occupied the time and attention of the U.S. National Library of Medicine over the past year.

Genetic Sequences and Public Access to Research Data

The European EMBL Nucleotide Sequence Database (also known as EMBL-Bank) and the U.S. GenBank both celebrated their 25th anniversaries in 2007. Possibly more significant was the 20th anniversary of the International Nucleotide Sequence Consortium, which connects these European and U.S. genetic sequence repositories with the DNA Data Bank of Japan in a highly successful collaboration to provide access to the ever increasing amount of genetic data produced by institutions around the world [1]. The Consortium provides a spectacular illustration of the value of standard electronic interchange formats and high speed telecommunications. The NLM's National Center for Biotechnology Information (NCBI) was instrumental in the development of the standards that now allow the three repositories to exchange newly received data every day, ensuring fast access to key scientific information across the globe. Responsibility for management of GenBank was transferred to NCBI from the Los Alamos National Laboratory in October 1992.

Sequence data provided a compelling early example of the power of journal editors to promote public deposit of scientific data, and of research funders to ensure that biomedical research information is available in public databases. By 1989, key journals

were requiring evidence of submission of sequence data to GenBank prior to publication of an article discussing a sequence [2]. A basic tenet of the funders of the Human Genome Project, launched in 1990, was that the data from all the international participants should be pooled and made available to researchers worldwide [3].

More recently the basic nucleotides of the original Human Genome sequencing project have been indexed to and complemented by corresponding (homologous) nucleotide sequences from genetic sites on more than 100,000 other species. All of these are directly linked where possible to published reports of studies that elucidate the substructure of the genes and the proteins produced. In many cases, these have now also been linked to the fine structure of the actual proteins coded for by the genes. Together these data represent a vast amount of human biology, most of which is not yet understood. For example, only two percent of the human genes so far identified are known to code for protein production and hence might have a known function. Opening these vast files of biological data to able enquiring minds worldwide speeds a more complete understanding of our species.

The success of the Human Genome Project amply demonstrated the scientific benefits of free access to pooled international research data, prompting the U.S. National Institutes of Health (NIH) and other international research funders to invest in public data resources. For example, the NIH Molecular Libraries project led to the NCBI's development of the PubChem database of small molecules and bioassay data in 2004. PubChem includes data deposited by companies and researchers from 13 countries and is accessed by as many as 40,000 users every day. More recently, NIH, the Office of the U.S.

Secretary of Health and Human Services, and private funders have joined together to support Genome Wide Association Studies, which link patients' clinical data to their genomic data to form a powerful new tool for identifying genetic and environmental factors that affect health, disease, and response to therapy. Funding is contingent on the resulting data being broadly available to the research community. Summary data are publicly available in dbGaP (database of Genome and Phenome), [4] another database developed by NCBI. Researchers may request individual patient level data (also stored in dbGaP) under a carefully crafted data-sharing plan [5] released in 2007. Clinical and genomic data from the monumental Framingham Heart study became available to researchers in dbGaP in October 2007. In addition to providing exciting new opportunities for biomedical research, dbGaP is also an important resource for studying clinical research protocols and identifying strategies for prospective alignment of clinical research variables with standards for electronic health care data.

Free MEDLINE and Public Access to Scientific Literature

2007 was also the 10th anniversary of free worldwide access to MEDLINE via PubMed, which - not incidentally - is linked to GenBank, other scientific databanks, full-text articles available on publisher's websites and in the PubMed Central electronic archive, libraries' electronic and print collections, and to consumer health information resources. NLM released an Internet-based interface to MEDLINE in 1996 and was able to eliminate its fees for searching MEDLINE and other databases in 1997 due to broad availability of Internet

access. As users began to pay for their own Internet connections for a variety of purposes, commercial telecommunications services were no longer needed to access MEDLINE, and NLM no longer had a need to charge users for their share of a large central telecommunications bill.

The Internet and the World Wide Web enabled NLM to make PubMed/MEDLINE free, but the strong impetus for doing so came from those attempting to provide information services to disadvantaged communities, as well as from biomedical researchers in relatively resource-rich environments. In 1994, prior to broad availability of the World Wide Web, NLM used an increase in its AIDS-related funding to make access to the AIDSLINE, AIDSDRUGS, and AIDSTRIALS, and DIRLINE databases free. This was in direct response to recommendations from community-based HIV/AIDS support organizations made at an NIH HIV/AIDS Information Services Conference [6] held in 1993. As an aid to the Human Genome project, NCBI began distributing MEDLINE citations to articles containing sequence data on CD-ROMs in 1991 and in 1994 provided free experimental online access via the Internet to this subset of MEDLINE. The value of free access to NLM's bibliographic data became obvious to the entire spectrum of user populations - and to the U.S. Congress - as NLM worked with the National Network of Libraries of Medicine in the U.S. to improve information services to underserved health professionals [7]. Use of personal computers and specialized interface software, such as Grateful Med® [8], made access to MEDLINE from offices and homes easier, but even the relatively modest fees charged for MEDLINE access by NLM, its International MEDLARS partners, and commercial database pro-

viders were prohibitive for many health professionals serving disadvantaged populations in the U.S. and elsewhere around the world. The various MEDLINE on CD-ROM products [9] produced by the private sector helped to provide more affordable access throughout the world - given local technical expertise and reliable electricity.

Once PubMed/MEDLINE was free via the World Wide Web, use immediately increased ten fold. It has been climbing steadily ever since. Since 1997, PubMed has been enhanced to include direct links from bibliographic citations to a range of scientific data, to consumer health information, and to the full text of articles on an expanding number of publisher websites through its "LinkOut" feature. PubMed has also been augmented to enable users to send electronic requests for articles to libraries and to allow libraries to provide customized "LinkOut" capabilities for their direct user populations (e.g., faculty and students in their universities) that facilitate access to the electronic (and print) journals in their collections [10]. These features improved access to biomedical information, but also highlighted the fact that many PubMed users were not able to view the electronic full-text of most articles due to their lack of affiliation with an institution with the necessary subscriptions or licenses.

Few biomedical journal articles were freely available on the Web in 1997. Research libraries - and the institutions they served - were faced with continuing journal subscription price increases they could ill afford - not to mention a confusing array of licensing arrangements for electronic journals, some with restrictions that seemed to counteract many of the long anticipated advantages of electronic publishing. Limits on access to and use of articles generated by their own faculty and students aroused particular frustration on university cam-

puses. The development of societal interest in "open source", "open access" and the public's right to electronic access to research results - and of research funding policies that promote public access to scientific articles - is a complex story with many significant events and fascinating characters. We refer readers to a recent book by Frances Groen [11] and to the web site of SPARC®, the Scholarly Publishing and Academic Resources Coalition [12] - itself a key player in these developments - for the wider context in which broader public access to biomedical journal articles has emerged.

In an environment increasingly receptive to significant change in scholarly communication patterns, scientific researchers - encouraged by the benefits of free Web access to molecular biology data and MEDLINE citations - began promoting the notion of an electronic full-text pre-print archive for biomedicine, similar to the physics archive initiated by Paul Ginsparg at Los Alamos National Laboratory in 1991 [13]. In 1999, Harold Varmus, then NIH Director, invited public comment on a proposed electronic archive of both published papers and pre-publication manuscripts to be built and managed by NCBI and linked with PubMed, GenBank, and other NLM databases. The proposal generated considerable controversy among publishers, editors, clinicians, and scientists [14,15] and also encouraged some biomedical publishers to make more articles freely available on their own websites. In response to feedback regarding the potential dangers of unreviewed clinical research papers, NIH's PubMed Central archive was released in 2000 with an initial scope limited to final published versions of articles voluntarily deposited by interested journal publishers. Then, as now, participating publishers entered into formal agree-

ments with NLM that might (or might not) specify a delay between initial publication of journal articles and their public availability in PubMed Central.

Journals depositing their current contents in PubMed Central could also elect to have their backfiles scanned and digitized by NLM for inclusion in the database. In 2004, the Wellcome Trust and the U.K. Joint Information Systems Committee provided substantial additional funding for scanning and digitizing the backfiles of selected British journals for addition to PubMed Central. The digitization of backfiles of journals for PubMed Central is just one example of a much broader international phenomenon. The Web has become an important vehicle for making newly digitized elements of the historical record of biomedicine and other fields readily available worldwide. This includes printed literature, rare books, manuscripts, images, and moving pictures. At the other end of the spectrum, many new reports, policy statements, press releases, etc. on biomedical and health topics are now published on the Web. The result is that very recent "gray literature" is more accessible than at any time in history, but perhaps also more in danger of disappearing tomorrow.

As the PubMed Central archive grew, there were other important efforts to promote international electronic access to biomedical journals. Especially notable is the Health InterNetwork Access to Research Initiative (HINARI), which provides free or very low cost online access to journals in biomedical and related social sciences to local, non-profit organizations in developing countries [16]. Organized under the auspices of the World Health Organization (WHO), HINARI was launched in 2002 with about 1,500 journals published by Blackwell, Elsevier Science, the Harcourt Worldwide STM Group, Wolters Kluwer International Health &

Science, Springer Verlag, and John Wiley. Twenty-two additional publishers joined later that year, bringing the total number of journals to over 2000. Organizations in developing countries apply for access through the WHO. Administrative and technical problems slowed start-up, and there have been recent reports of reduced access to important journals previously available [17]. Nevertheless, HINARI has improved access to biomedical literature in many countries. NLM worked with HINARI's technical management to establish it as a "LinkOut" provider so that HINARI members can connect directly to full-text articles from PubMed.

As the number of electronic journals increased so did interest in the development of robust standards for them. Publishers sought standard or recommended formats that would ease their transition to electronic publishing. Libraries concerned with permanent access to electronic content knew that standard formats would simplify the enormous job of digital preservation for this category of material. Given its experience with ingesting electronic journals from multiple sources and converting them to a single tagged format for PubMed Central, the NCBI had both experience and strong motivation to develop a standard format that would accommodate all life sciences journals. As NCBI was developing the format, the Harvard University Libraries, with funding from the Mellon Foundation, provided additional resources for analysis of journals so the format would cover all disciplines. The first version of the NLM Journal Archiving and Interchange DTD Suite [18] appeared in 2003. An international Working Group advises on necessary changes and extensions to the Suite, which now covers books as well as journal articles. In 2006, the U.S. Library of Congress and the British Library jointly endorsed the

NLM DTD as the standard for digital preservation of electronic journals. It is now a de facto international standard.

PubMed Central was an enabler for biomedical research funders wishing to ensure that articles arising from supported research were readily available to researchers, students, and the general public. Encouraged by strong interest in the U.S. Congress, in 2004 NIH Director Elias Zerhouni issued a proposed policy for building a comprehensive archive of NIH-funded research to enable more effective and efficient research portfolio management and to ensure broad public access to articles arising from NIH-funded research. The draft policy called for mandatory submission to PubMed Central of peer-reviewed final manuscripts reporting NIH-funded research for public release within 6 months of publication. In response to strong reactions from both commercial and society publishers, who predicted dire consequences for the sustainability of scholarly communication and of scientific associations, the final policy effective in May 2005 strongly encouraged voluntary submission of the manuscripts for public release within 12 months of publication. NCBI built an efficient individual manuscript submission system to implement the NIH policy. The system was also used to implement a mandatory 6-month policy established by the Wellcome trust and similar research policies by other funders in the United Kingdom and elsewhere. NCBI also set up batch pipelines for publishers who wished to submit final manuscripts (which required final review by authors after tagging for inclusion in PubMed Central) or final published papers on behalf of NIH-funded authors. The voluntary NIH public access policy did not achieve its desired goal. Only 5 percent of applicable articles were submitted by NIH-funded authors, and

submissions of such articles through all mechanisms (including routine publisher deposit of journals in PubMed Central) never reached 25 percent of relevant papers. In the meantime, the mandatory policies of other research funders had no noticeable negative effect on biomedical publishers. Dissatisfied with the rate of progress on public access to NIH-funded articles, the U.S. Congress included a mandatory public access provision in the Appropriations Act for fiscal 2008, which became law on December 26, 2007. On January 11, 2008, the NIH issued implementation guidance for mandatory submissions which becomes effective on April 7, 2008 [19].

NLM computer systems are now answering close to two billion inquiries per year. Biomedical knowledge arises worldwide. Even in the first 1879 volume of *Index Medicus*, the majority of citations were to non-U.S. articles. Allied computer facilities outside the U.S. clearly are now needed to provide local access and back-up for PubMed Central holdings. In response to inquiries from a number of countries, the NCBI developed PubMed Central International (PMCI), a lighter weight version of the system that depends on the NLM standard DTD. Once again, the Wellcome Trust took a leadership role and set up a partnership of public and private research funders to establish UKPMC in January 2007 as the first operational PubMed Central International site. UKPMC currently serves as a mirror site for the majority of the content in NIH's PubMed Central site and also ingests manuscripts supported by UK biomedical research funders. Institutions in the Europe, the Far East, and Africa are also considering setting up PubMed Central International sites.

Just 10 years after MEDLINE became free, 46 percent of the 17.8 million bibliographic citations in PubMed

are linked to electronic full-text articles. For the 4.1 million articles published from 2000 on, the percentage linked to full-text is 84 percent. The percentages linked to free full text are much lower: 12 percent for the entire file and 19 percent for 2000 on - but all of these numbers are climbing gradually. Retrospective digitization of journals and of pre-MEDLARS indexing data (now back through 1949) has greatly enhanced access to historical literature. Just 7 years after it was established, PubMed Central contains more than 1.4 million articles. More than 3 million users retrieve more than 12 million copies of articles from it every month.

Consumer Health Information and Public Access to Clinical Trials Data

With the birth of the Web browser in 1993, the Internet became an obvious vehicle for improving access to health information written for patients, their families, and the general public. Interesting health sites immediately appeared, e.g., drkoop.com, but sustainable business models lagged behind. Internet users did not expect to pay for information, and Web advertising had yet to take off. NLM's entry into the field of Web-based health information for the public was another development prompted by free MEDLINE access. Once MEDLINE was free, large numbers of the lay public began to use it, even though it was scarcely the best place to attain the basic understanding of health and disease topics that many of them were obviously seeking. The doubling of the NIH budget gave NLM the opportunity to develop new services targeted toward the general public. In 1998, NLM released the MedlinePlus consumer health information website

[20], with web information on 22 frequently searched disease topics from NIH and other authoritative sources, carefully organized and evaluated by medical librarians. With no advertising and no registration requirements, today MedlinePlus has more than 700 health topics in both English and Spanish, as well as health news, interactive tutorials, dictionaries, an illustrated encyclopedia, drug and dietary supplement information, videos of surgical operations, and specific links to NLM databases, such as PubMed, ClinicalTrials.gov, toxicological resources, and Genetics Home Reference. The "Go Local" feature connects authoritative health information in MedlinePlus to sites that provide information about health services available in many localities within the U.S. Teams of librarians at NLM and at Go Local sites review content regularly to ensure its continued currency and appropriateness. MedlinePlus delivers 250 million page views to close to 100 million users each quarter. About 42 percent of the use is from outside the U.S.

The development of ClinicalTrials.gov, [21] now an international registry of clinical studies, was precipitated by another event in 1997: the passage of the Food and Drug Administration (FDA) Modernization Act. This law instructed the U.S. Secretary of Health and Human Services to establish a registry of trials of FDA-regulated drugs intended to treat "serious or life threatening conditions" - in consultation with the administrator of FDA and the directors of NIH, NLM, and the Centers for Disease Control and Prevention. The purpose of the database was to expand the U.S. public's access to ongoing trials that might be relevant to them. NLM was assigned responsibility for the database, which was launched in 2000 with NIH-funded trials and later expanded to include other federally-funded trials and industry trials. There

were no penalties for non-compliance with registration under FDAMA, and less than 10,000 trials were registered.

In 2004, the power of journal editors to compel submission of research data was once again demonstrated when the International Committee of Medical Journal Editors (ICMJE) announced that effective with trials beginning on July 1, 2005 or later they would refuse to publish the results of clinical trials that were not registered in a public trial registry prior to the enrollment of the first patient [22]. Ongoing trials were given until September 13, 2005 to register. At the time of the ICMJE announcement, ClinicalTrials.gov was the only public registry that met their requirements. In response to the ICMJE action, NLM opened up the database to international trials without a U.S. site. The ICMJE action was prompted by some celebrated cases in which trial results were incompletely reported in journal articles, with the result that serious adverse effects affecting some populations were essentially concealed from the public. Journal editors required trial registration so they could compare results reported in submitted papers to registry entries to detect incomplete reporting, discrepancies in outcome measures, etc. The ICMJE policy prompted a sharp increase in the number of trials registered in ClinicalTrials.gov [23] and the number of organizations registering, both inside and outside the U.S., making the database more valuable to the public, journal editors, and those engaged in systematic review of evidence. Reflecting the strong international interest in registration, in 2005 the World Health Organization (WHO) initiated a project to develop standards for international trial registries [24]. This effort was welcome - and ClinicalTrials.gov includes the WHO standard elements - but WHO's expressed interest in the estab-

lishment of national clinical trial registries was less helpful. A careful collaboration among a few large regional registries - following the model of the International Nucleotide Sequence Consortium - would be a more cost-effective and sustainable approach that would reduce duplicate registrations of multinational trials.

Ten years after action by the U.S. Congress prompted development of the ClinicalTrials.gov, the recent FDA Amendments Act of 2007 has expanded mandatory registration to include additional types of trials (devices as well as drugs) and more data elements (essentially all those required by the ICMJE guidelines), mandated deposit of unpublished results data in a public database, and imposed significant penalties for non-compliance. NLM is now engaged in expanding ClinicalTrials.gov to meet these challenging new requirements. Expanded registration was effective on December 26, 2007, just 90 days after the new law was passed. The new mandate prompted another significant surge in submissions to ClinicalTrials.gov both from the U.S. and from other countries, including thousands probably not subject to the new law. The database currently contains more than 51,000 trials from 153 countries. Public access to clinical trial results that have not been peer-reviewed for publication is uncharted territory - both on a technical and public policy front [25]. Considerable experimentation is likely to be required to achieve presentation methods that provide appropriate context for clinicians and the public.

Open Access to Image Data and Software Tools

Just as high speed computing and telecommunications are essential to effec-

tive use of molecular biology and genomic data, they are a pre-requisite for serious use of high resolution image data. NLM began the Visible Human project [26] to produce a public domain three-dimensional database of cryographic, magnetic resonance imaging, and computed tomography images of a complete human male and female at sub-millimeter resolution before desktop computers were capable of handling files of that magnitude. By the time the Visible Human data sets became available for FTP download over the Internet and on CD-ROMs in 1994 and 1995, many researchers and system developers had local systems capable of handling them. Five years ago, NLM funded a group of universities and companies to jointly develop and test the Insight Toolkit (ITk) [27], an open source software toolkit for segmentation and registration of tomographic image data sets, including those in the Visible Human data sets. The goal was to create a dynamic, public domain and extensible toolkit usable in imaging research and application development throughout the world. Public availability of the datasets and the toolkit have had a major impact on imaging research worldwide and led to significant applications in radiological imaging, educational simulations, and surgical planning [28].

Ongoing Challenges

The combination of the Internet, Google®, other search engines, and ever-expanding stores of free information in electronic form has improved access to biomedical knowledge immeasurably since the first Web browser was released less than 15 years ago. Unfortunately, increasing the amount of electronic biomedical information that is readily available can actually decrease

the chances of finding the most relevant information quickly. Linking information intelligently from disparate sources and delivering the right information in time to improve decision-making for a wide range of users are therefore perpetual goals - with informatics research challenges that change as new sources of information and new technologies become available. Virtually every information resource described in this article makes use of the Unified Medical Language System (UMLS) [29] resources in some way to compensate for variations in the ways concepts are expressed; most of them are richly linked to other information sources (including some patient records) to help users discover and connect to the information most relevant to their inquiries; and many are frequent targets of natural language processing, data-mining, and automated analysis efforts. Although significant advances have been made, we are a long way from systems that behave as if they really know what users want or need - and then give it to them quickly.

We are also at the dawn of a new form of electronic publication that will carry special access, use, and preservation challenges. Interactive publications, while mostly a future development, have already been sighted at NLM [30]. In an analysis done at the Library in 2005, about 2 percent of the articles in sample of likely journals indexed in MEDLINE contained electronic elements designed to respond to on-line user actions and the number of such articles is increasing. Some articles include detailed depictions of surgical findings and procedures, even demonstrations of the underlying embryological abnormalities. In other cases the articles include "the data behind the tables". What potential, if any, do such publications have for speeding understanding of new knowledge and pro-

longing its effective retention? NLM has entered into partnership with the Optical Society of America to determine the most effective examples of interactive publication, their costs, and possible means by which NLM could in the future preserve such digital publications.

No information is useful if it is not accessible when needed. Natural and man made disasters in the past decade have repeatedly shown seemingly silly, but important faults in disaster information management that are easily understandable by the public as well as by medical informaticians, e.g., police and fire service telecommunication equipment broadcasting on incompatible frequencies, huge numbers of evacuees without knowledge (let alone electronic records) of their medications. In addition, the Katrina hurricane disaster in Louisiana/Mississippi demonstrated the difficulty in helping thousands of old, feeble residents of nursing homes - evacuated to public shelters - who could not identify themselves.

Possible solutions to these problems are under investigation within NLM's Lister Hill Center and the Library's Division of Toxicology and Environmental Health, and by participating libraries in the National Network of Libraries of Medicine. As demonstrated by the Department of Veterans Affairs system and the rapid pooling of medication data from pharmacy chains during the Katrina disaster, electronic medical records can be used to help displaced persons during large-scale disasters. Personal health records, whether portable or remotely accessible, could be similarly beneficial. NLM is exploring approaches to making it easy for individuals who create personal health records to incorporate standard terminology and identifiers in them. The Library continues to support broad availability of standard clinical terminologies as one prerequisite for

electronic health records that can connect effectively to each other and to electronic biomedical knowledge, whether in routine health care or in the aftermath of disasters. NLM represented the U.S. in the nine-country negotiations that led to the formation of the International Health Terminology Standards Development Organization (IHTSDO) in 2007 and its acquisition of the SNOMED CT clinical terminology from the College of American Pathologists [31]. This may prove to be an important step toward useful international standardization of health data.

Electronic health records are an old idea with new relevance to disaster information management. Newer ideas such as implantable electronic identity devices (such as the RFIDs used in pets), while not without ethical issues, are under serious consideration in the US [32]. If adopted, national registries designed to preserve personal health identities would provide at least one emergency path to critical health information. The National Network of Libraries of Medicine is testing the expansion of MedlinePlus Go Local coverage to include resources required in disasters and emergencies, as well as various approaches to integrating libraries into community disaster preparedness and response efforts.

As articulated in the NLM Board of Regents Long Range Plan for 2006-2016 [33], the Library's first responsibility is to ensure effective uninterrupted access to its own heavily used electronic sources of biomedical information both in times of disaster or emergency and in response to changing technologies and user needs. Organizations, including universities, which have or seek a major mandate to build and maintain public repositories of knowledge in the digital age must recognize and plan for the substantial added costs of data acquisition, quality

control, organization, and preservation, plus the responsibility to provide up to date access software and hosting. Those universities that do not seek or accept such a mandate will perhaps limit their scope to arranging access to information for their faculty and students and possibly to storage and preservation of information generated within their institutions.

Many organizations are currently heavily dependent upon the NLM and other governmental institutions around the world for maintenance of public scientific data bases and other biomedical information resources. Although he sees biomedicine as a potential exception, Clifford Lynch, Executive Director of the Coalition for Networked Information has stated that "ultimate responsibility for stewardship in many disciplines, after the vagaries of government funding over time are considered, will have to rest with higher education rather than the funders of scholarly and scientific research." [34] This is a very important question. From NLM's viewpoint, intellectual fervor and creativity still abound in universities and in the biotechnology, biomedical engineering and pharmaceutical industries, too, but we cannot urge societal reliance upon these sectors to provide permanent reliable public access to quality assured - let alone free - biomedical knowledge. This has seemed a proper function to be assayed by governments. Lack of monopoly on wisdom by any segment of society suggests that to some extent parallel efforts continue to make sense. Threats of physical and political disasters suggest that duplicate stores of key scientific information on two or three continents may continue to be prudent and practical - just as they proved essential to the preservation of knowledge during the wars and disasters of the 20th and many previous centuries.

References

- Public Collections of DNA and RNA Sequence Reach 100 Gigabases [press release on the Internet]. Bethesda, MD: National Library of Medicine, 2005 Aug 22 [cited 2008 Feb 27]. Available from: http://www.nlm.nih.gov/news/press_releases/dna_ma_100_gig.html.
- Dawid IB. Submission of sequences. *Proc Natl Acad Sci USA* 1989 January;86(2):407.
- Watson JD, Cook-Deegan RM. Origins of the Human Genome Project. *FASEB J* 1991 Jan;5(1):8-11.
- Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007 Oct;39(10):1181-6.
- National Institutes of Health. Genome Wide Association Studies. Guidance for Developing Data Sharing Plans for GWAS Studies [Internet]. Bethesda, MD: National Institutes of Health, 2007 Nov. [cited 2008 Feb 27]. Available from: http://grants.nih.gov/grants/gwas/gwas_data_sharing_plan.pdf.
- National Library of Medicine. Information services for HIV/AIDS: recommendations to the NIH. report of a conference co-sponsored by the National Library of Medicine and NIH office of AIDS research, June 28-30, 1993. [Internet]. NIH publication no. 94-3730, 1994. [cited 2008 Feb 27]. <http://sis.nlm.nih.gov/hiv/nihprt.pdf>.
- Wallingford KT, Ruffin AB, Ginter KA, Spann ML, Johnson FE, Dutcher GA, et al. Outreach activities of the National Library of Medicine: a five-year review. *Bull Med Libr Assoc* 1996 Apr;84(2 Suppl):1-60.
- Haynes RB, McKibbin KA. *Grateful Med*. MD Comput 1987 Sep-Oct;4(5):47-9, 57.
- Woodsmall RM, Lyon-Hartmann B, Siegel ER, editors. *MEDLINE on CD-ROM: National Library of Medicine evaluation forum*, Bethesda, Maryland, September 23, 1988. Medford, NJ, USA: Learned Information; 1989.
- Schott MJ. PubMed enhancements: fulfilling the promise of a great product. *Med Ref Serv Q* 2004 Winter;23(4):1-11.
- Groen FK. *Access to medical knowledge: libraries, digitization, and the public good*. Lanham, MD: Scarecrow Press; 2007.
- SPARC, The Scholarly Publishing & Academic Resources Coalition [Internet]. Washington, D.C.: SPARC, c2007 [cited 2008 Feb 27]. Available from: <http://www.arl.org/sparc/>.
- Ginsparg P. As we may read. *J Neurosci* 2006 Sep 20;26(38):9606-8.
- Fox R. E-biomed. *J R Soc Med* 1999 Sep;92(9):441-2.
- Relman AS. The NIH "E-biomed" proposal-a potential threat to the evaluation and orderly dissemination of new clinical studies. *N Engl J Med* 1999 Jun 10;340(23):1828-9.
- Katikireddi SV. HINARI: bridging the global information divide. *BMJ* 2004 May 15;328(7449):1190-3.
- Villafuerte-Gálvez J, Curioso WH, Gayoso O. Biomedical journals and global poverty: is HINARI a step backwards? *PLoS Med* 2007 Jun;4(6):e220.
- National Library of Medicine. NLM Journal Archiving and Interchange Tag Suite. [Internet]. Bethesda, MD: National Library of Medicine, 2007 Nov 1 [cited 2008 Feb 27]. Available from: <http://dtd.nlm.nih.gov/>.
- National Institutes of Health. Revised Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research [Internet]. Bethesda, MD: National Institutes of Health, 2008 Jan 11 [cited 2008 Feb 27]. Available from: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-033.html>.
- Miller N, Lacroix EM, Backus JE. MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service. *Bull Med Libr Assoc* 2000 Jan;88(1):11-7.
- McCray AT. Better access to information about clinical trials. *Ann Intern Med* 2000 Oct 17;133(8):609-14.
- De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, Koet al.; International Committee of Medical Journal Editors. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *CMAJ* 2004 Sep 14;171(6):606-7.
- Gülmezoglu AM, Pang T, Horton R, Dickersin K. WHO facilitates international collaboration in setting standards for clinical trial registration. *Lancet* 2005 May 28-Jun 3;365(9474):1829-31.
- Zarin DA, Tse T, Ide NC. Trial Registration at ClinicalTrials.gov between May and October 2005. *N Engl J Med* 2005 Dec 29;353(26):2779-87.
- Zarin DA, Ide NC, Tse T, Harlan WR, West JC, Lindberg DA. Issues in the registration of clinical trials. *JAMA* 2007 May 16;297(19):2112-20.
- Ackerman MJ. The Visible Human Project. *J Biocommun* 1991;18(2):14.
- Yoo TS, Metaxas DN. Open science-combining open data and open source software: medical image analysis with the Insight Toolkit. *Med Image Anal* 2005 Dec;9(6):503-6.
- Burke C, Patrias K, compilers. Visible Human Project [bibliography on the Internet]. Bethesda (MD): National Library of Medicine (US); 2007 May [cited 2008 Mar 08]. (Current bibliographies in medicine; 2007-1). 912 citations from January 1987 through March 2007. Available from: <http://www.nlm.nih.gov/pubs/cbm/visiblehuman2007.html>.
- Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. *Methods Inf Med* 1993 Aug;32(4):281-91.
- Lindberg DA. Research opportunities and challenges in 2005. *Methods Inf Med* 2005;44(4):483-6.
- International Health Terminology Standards Development Organisation. [Internet] Copenhagen: IHTSDO, c2008 [cited 2008 Mar 08]. Available from: <http://www.ihtsdo.org>.
- Anderson AM, Labay V. Ethical considerations and proposed guidelines for the use of radio frequency identification: especially concerning its use for promoting public safety and national security. *Sci Eng Ethics* 2006 Apr;12(2):265-72.
- National Library of Medicine. Board of Regents. Charting a course for the 21st century: NLM's long range plan 2006-2016. Bethesda, MD: National Library of Medicine; 2006.
- Lynch CA. A matter of mission: information technology and the future of higher education. In: Katz R. editor. *The tower and the cloud*. Boulder, CO: EDUCAUSE. In press 2008.

Correspondence to:

Donald A. B. Lindberg, M.D.
National Library of Medicine
8600 Rockville Pike
Bethesda, MD, 20894
USA
E-mail: lindberg@nlm.nih.gov