

## Preface - Access to Knowledge Revisited

Donald A.B. Lindberg<sup>1</sup>, Betsy L. Humphreys<sup>2</sup>

<sup>1</sup> Director Emeritus, National Library of Medicine, Bethesda, Maryland, USA

<sup>2</sup> Acting Director, National Library of Medicine, Bethesda, Maryland, USA

### Summary

**Objective:** To review and update the Preface to the 1998 Yearbook of Medical Informatics, which had as its Special Topic "Health Informatics and the Internet".

**Method:** Assessment of the accuracy of predictions made in 1998 and consideration of key developments in informatics since that time.

**Results:** Predictions made in 1998 were generally accurate regarding reduced dependence on keyboards, expansion of multimedia, medical data privacy policy development, impact of molecular biology on knowledge and treatment of neoplasms, and use of imaging and informatics to advance understanding of brain structure and function. Key developments since 1998 include the huge increase in publicly available electronic information; acknowledgement by leaders in government and science of the importance of biomedical informatics to societal goals for health, health care, and scientific discovery; the influence of the public in promoting clinical research transparency and free access to government-funded research results; the long-awaited arrival of electronic health records; and the "Cloud" as a 21st century reformulation of contracting out the computer center.

**Conclusions:** There are many challenging and important problems that deserve the attention of the informatics community. Informatics researchers will be best served by embracing a very broad definition of medical informatics and by promoting public understanding of the field.

### Keywords

Access to information, consumer health information, informatics, electronic health records, brain mapping, neoplasm gene

Yearb Med Inform 2016;Suppl1:S18-20

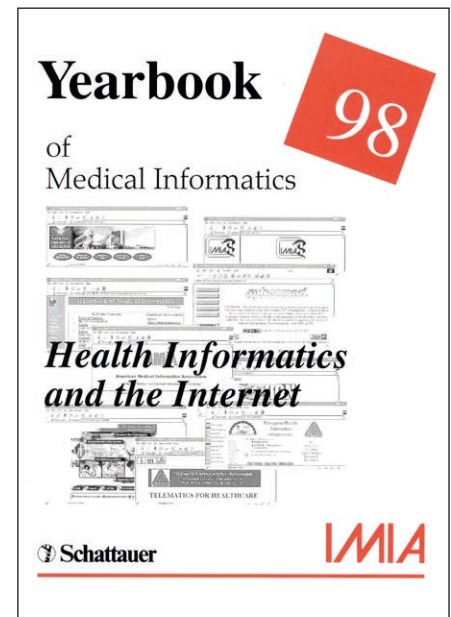
<http://dx.doi.org/10.15265/IYS-2016-s026>

Published online May 20, 2016

In 1998, the Special Topic for the *Yearbook of Medical Informatics* was "Health Informatics and the Internet", acknowledging the transformational impact on medical informatics of developments in worldwide network connectivity emanating from outside the field.

*Access to Knowledge* was an obvious choice for a Preface written by the Director of the U.S. National Library of Medicine (NLM) [1] less than a year after MEDLINE became freely available worldwide on the Internet via PubMed. The spread of the Internet, the World Wide Web, and Web browsers allowed NLM to end its reliance on commercial telecommunication networks (thus eliminating the basis for user fees), to cease the distribution of search software on physical media for multiple hardware platforms, and to link MEDLINE citations to related information, including the electronic full text of journal articles and sequence data in GenBank. There was an immediate 10-fold increase in searches on the NLM system once MEDLINE was free via the Internet, as much as 30% by patients, families, and the public.

The 1998 preface contemplated the potential impact of free MEDLINE on the long-predicted transition from paper to electronic journals. As reported in the *Yearbook* in 2008 [2], free MEDLINE did play a supporting role in the subsequent rapid move to electronic journals in medicine and science and also in the promotion of public/open access to electronic journals and the notion that people are entitled to free access to the results of research funded by their taxes. In 1998, celebration of free public access to scientific and medical information in MEDLINE was tempered by a caution that greater availability of information does not



automatically provide patients, families, and the public with "understanding, discernment, or sophistication of choice". This led NLM to develop information services specially designed to help the public improve its understanding of health topics [3]. Caution is even more valid today, for clinicians as well as for the public, given the exponential increase in the amount and range of information freely available in multiple formats and via multiple channels.

The predictions made in the 1998 preface turned out to be for the most part accurate. As envisioned, control of technology by speech, touch, and gesture has indeed reduced dependence on keyboards, and retrieved information increasingly combines images, text, and speech in useful ways. Not anticipated was the extent to which social media, smartphones, wearable technology, and the "Internet of Things" would influence these advances.

Again as predicted, there has been modest progress in legislation, regulation, and research policy related to medical data privacy, which does indeed remain highly important to progress in informatics and genetics. Viable, if imperfect, approaches to blinding the identity of individuals represented in health care and clinical research databases,

and to implementing rational access controls have dramatically increased the aggregated clinical data available for research [e.g., 4, 5]. One problem mentioned in 1998, i.e., encryption schemes that are practical, and strong, and can coexist at some level with national and international security protections, is not, and may never be, solved. Since 1998, we have learned that exercising due diligence to deflect cybersecurity attacks and prevent data breaches is a major, endless, and increasingly expensive task for every organization with a health data system [6].

Expected to arrive perhaps sooner than it did, the marriage of molecular biology and informatics has improved understanding of neoplasms and methods of attacking them, forming the basis for the cancer component of the current major U.S. Precision Medicine Initiative [7]. Also as predicted, advances in imaging and informatics are leading toward increased understanding of brain structure and function [8].

Beyond the obvious huge increase in the amount of electronic knowledge, information, and data available to the informatics field and to the public, what is most remarkable about the developments of the past 18 years? Here are a few candidates:

- Leaders in government and scientific research now see the central importance of biomedical informatics to broader societal goals for health, health care, and discovery of new knowledge. The language is different, e.g., Big Data, Data Science, Data Mining, Health Data Interoperability, Precision Medicine, but the substance is familiar. Computer technology is now essential for all modern biological science, as well as much modern engineering, physics, and mathematics, and this is understood and accepted. Comprehension of its importance to clinical research is on the rise.
- The power of the people – amplified by ready access to inexpensive computing devices, the World Wide Web, and social media – has played a key role in expanding free access to biomedical information and data, advancing the concept of patients as engaged partners in health care that is centered around *them*, and promoting the view of people as participants in – as opposed to subjects of – clinical

research. Public concern or outrage was a strong force in prompting widespread requirements for public access to the published results of research funded by governments, public registration of clinical trials, and public availability of summary trial results. Despite lagging compliance enforcement, ClinicalTrials.gov has grown exponentially since its debut in 2000 and is now a novel resource for research on the clinical research enterprise itself, as well as a tool in clinical knowledge discovery [e.g., 9, 10]. The debate has now moved on to public access to individual participant data [11]. Patients increasingly have access to at least some of the information in their own electronic health records (EHRs) and have begun to contribute data to these records, with a reasonable prospect of more to come. Members of the public were active partners in planning the 1 million+ person cohort that will be the centerpiece of the U.S. President's Precision Medicine Initiative [12].

- EHRs have finally arrived – even in the U.S. They aren't perfect, as is evident from loud complaints about the lack of usability and interoperability of the available commercial EHR products. Current products do, however, have one absolutely essential prerequisite for real and continuing improvement: they are being broadly used. Legislation and regulation are blunt and imperfect instruments for advancing use of health information technology and related standards. Nonetheless, the U.S. HITECH Act and related "Meaningful Use" and EHR certification regulations have forced a level of use of EHRs and specific standards that has prompted salutary actions for improvement of both – in the U.S. and elsewhere [e.g., 13, 14]. Although there are glimmers of hope, we don't yet have good methods for summarizing the key information about the current patient, widespread integrated decision support for clinicians and patients, or a "Learning Health System" [15]. What we do have is a firmer foundation on which to build systems that are more intelligent, more capable, and easier to use; some feasible ideas for moving forward [16];

and progress on infrastructures that will enable clinicians to consider a patient's genetic make-up when making treatment decisions [17, 18, 19].

- The old idea of contracting out the computer center has re-emerged in the more flexible shape of the "Cloud". Although not without pitfalls, it can help to enable flexible on-demand use of advanced computing to analyze very "Big" biomedical data without robust local computing, download, and storage capacity [20].

Balancing out these largely positive developments is a lack of significant progress in other arenas.

Computing hardware is seemingly frozen in a physics Neverland awaiting the essential next step: optical computing. Major advances with current technology are blocked by the need to dissipate the heat associated with computing and the local retail cost of electric power. These are important engineering considerations that are unrelated to any truly new or different information needs. On the software side, there is still a dearth of products that apply and take effective advantage of massively parallel computing power, and of products with human-computer interfaces that satisfy normal user expectations.

An exception is the IBM Watson approach [21]. IBM achieved a milestone in artificial intelligence research, combining massively parallel computing with a heavy reliance on computational linguistic techniques for understanding natural language utterances. Remarkably, the computers used were ordinary commercial machines and the critical systems software was open source.

On the purely medical scene, despite nearly ubiquitous access to increasing stores of free high quality medical information and the emergence of commercial products with good summarized evidence relevant to many clinical decisions, we have not succeeded in ensuring that all those who deliver care in advanced medical treatment facilities have immediate and suitable access to the latest and best evidence. Like many aspects of health care, this is a matter that requires dedicated resources and continuing attention from multidisciplinary teams, including information specialists. Patient populations, scientific evidence, EHRs, personal

computing devices, health care workflows, and availability and cost-effectiveness of relevant information sources all evolve at different rates. The infrastructure to ensure that clinicians have efficient access to current knowledge must evolve, too. It won't happen by accident.

As even this very incomplete summary shows, there are many challenging and important problems that deserve serious thought, study, and action by informaticians. A very broad definition of medical informatics will serve the field best. This should certainly include computer-aided learning systems for the public, as well as for health professionals, and advanced decision support for difficult life choices, e.g., end of life care, participation in clinical trials, privacy trade-offs. We should court the public interest. Broader understanding of this field – its lofty goals as well as its current stumbling blocks – can only be of help to us.

### Acknowledgement

Supported in part by the Intramural Program of the NIH National Library of Medicine.

### References

- Lindberg, DA. Preface – Access to Knowledge. *Yearb Med Inform* 1998:1-3.
- Lindberg DA, Humphreys BL. Rising expectations: access to biomedical information. *Yearb Med Inform* 2008:165-72.
- 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.
- 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.
- Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform* 2015;216:574-8.
- Halamka, JD. The State of Information Security 2015. *Life as a Healthcare CIO*. 2015 Dec 9 [cited 2016 Jan 9] Available from <http://geekdoctor.blogspot.com/2015/12/the-state-of-information-security-2015.html>
- Zhao Y, Polley EC, Li MC, Lih CJ, Palmisano A, Sims DJ, et al. GeneMed: An Informatics Hub for the Coordination of Next-Generation Sequencing Studies that Support Precision Oncology Clinical Trials. *Cancer Inform* 2015 Mar 19;14(Suppl 2):45-55.
- Jorgenson LA, Newsome WT, Anderson DJ, Bargmann CI, Brown EN, Deisseroth K, et al. The BRAIN Initiative: developing technology to catalyze neuroscience discovery. *Philos Trans R Soc Lond B Biol Sci* 2015 May 19;370(1668).
- Califf RM, Zarin DA, Kramer JM, Sherman RE, Aberle LH, Tasneem A. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007-2010. *JAMA* 2012 May 2;307(17):1838-47.
- Wu M, Sirota M, Butte AJ, Chen B. Characteristics of drug combination therapy in oncology by analyzing clinical trial data on ClinicalTrials.gov. *Pac Symp Biocomput* 2015:68-79.
- Drazen JM. Sharing individual patient data from clinical trials. *N Engl J Med* 2015;372:201–2.
- Advisory Committee to the Director, National Institutes of Health. Precision Medicine Initiative (PMI) Working Group. The Precision Medicine Initiative Cohort Program – Building a Research Foundation for 21st Century Medicine. 2015 Sept 17. [Cited 2016 Jan 9] Available from [http://acd.od.nih.gov/reports/PMI\\_WG\\_report\\_2015-09-17-Final.pdf](http://acd.od.nih.gov/reports/PMI_WG_report_2015-09-17-Final.pdf)
- HL7 launches Argonaut Project to advance FHIR interoperability standard. *Health Manag Technol* 2015 Feb;36(2):26.
- Joint Initiative on SDO Global Health Informatics Standardization. [Cited 2016 Jan 9] Available from <http://www.jointinitiativecouncil.org/>.
- Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc* 2015 Jan;22(1):43-50.
- Mandl KD, Mandel JC, Kohane IS. Driving Innovation in Health Systems through an Apps-Based Information Economy. *Cell Syst* 2015 Jul;1(1):8-13.
- MacArthur JA, Morales J, Tully RE, Astashyn A, Gil L, Bruford EA, et al. Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. *Nucleic Acids Res* 2014 Jan;42(Database issue):D873-8
- Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016 Jan 4;44(D1):D862-8.
- Rubinstein WS, Maglott DR, Lee JM, Kattman BL, Malheiro AJ, Ovetsky M, et al. The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. *Nucleic Acids Res* 2013 Jan;41(Database issue):D925-35.
- Feltus FA, Breen JR 3rd, Deng J, IZard RS, Konger CA, Ligon WB 3rd, et al. The Widening Gulf between Genomics Data Generation and Consumption: A Practical Guide to Big Data Transfer Technology. *Bioinform Biol Insights* 2015 Sep 23;9(Suppl 1):9-19.
- Ferucci DA. Introduction to “This is Watson”. *IBM Journal of Research and Development* 2012 May-Jul;3-4:1-15.

### Correspondence to:

Donald A.B. Lindberg, MD  
13601 Esworthy Road  
Germantown, MD 20874-3319  
USA  
E-mail: donald.lindberg@gmail.com