The Next Generation Internet and Health Care: A Civics Lesson for the Informatics Community

Edward H. Shortliffe, MD, PhD
Stanford University School of Medicine
Stanford, California

The Internet provides one of the most compelling examples of the way in which government research investments can, in time, lead to innovations of broad social and economic impact. This paper reviews the history of the Internet's evolution, emphasizing in particular its relationship to medical informatics and to the nation's health-care system. Current national research programs are summarized and the need for more involvement by the informatics community and by federal health-care agencies is emphasized.

INTRODUCTION

Although the Internet may at first seem to be an overnight phenomenon[1], the history of its development and introduction goes back over 30 years. It is instructive to learn about its evolution, and about the gradual way in which it has penetrated not only our culture but also our thinking about scientific research and health-care delivery.

The medical informatics community has been involved as users and experimenters almost from the Internet's beginning. If we are to understand what lies ahead, and how to achieve the Net's potential, we need to understand how we achieved what we have today as well as the missed opportunities and the nature of the barriers that still exist.

In this paper I will offer a summary of the evolution of the Internet, emphasizing a biomedical perspective. I will also summarize recent organizational and logistical developments, propose some likely future directions, and provide my views on the role that the health-care community could and should be playing as the technology evolves. Much of what follows is inevitably colored by personal recollection and opinion.

HISTORICAL PERSPECTIVE

The technology of packet-switched networking arose in the 1960s, and it was in the latter half of that decade that the Department of Defense, through its Advanced Research Projects Agency (ARPA), sought to use the technology to link together a handful of computers that were involved with defense-related research[2]. Some of these machines were on university campuses, while others were at federal sites or in the facilities of government contractors. Recall that this was the era of an unpopular war in Southeast Asia, and there was much suspicion on college campuses about the motives behind this kind of technology and its potential military uses.

By the 1970s, however, it became clear that the ARPANET, as this network became known, was a boon to collaborative research in computer science. Although the initial emphasis of the network had been on remote login to computers (Telnet) and file sharing among machines (File Transfer Protocol, or FTP), an early application known as electronic mail was a surprise hit. Email quickly penetrated the ARPANET research community and accounted for much of the traffic on the national network. In addition, by the late 1970s Ethernet technology had been introduced and the first local area networks were being implemented on campuses. These facilitated connectivity to the national networks from various campus locations. By 1982 the networking protocol known as TCP/IP had been introduced; it became the dominant protocol both on the national network and, in time, on local networks as well.

As additional universities and research institutes became connected to the ARPANET, the need arose for more robust addressing conventions than the early single-machine naming approach. Several new naming systems were introduced until the domain system of today (with the familiar .edu, .org, .gov, .com, and .net suffixes) was eventually implemented. Network speeds increased, and an ARPANET culture began to evolve. There was a strong sense of community, of openness and free speech, and of the need to avoid commercial activities. As recently as the late 1980s, there was still no consensus whether commercial organizations should be allowed to connect to the Net (other than government-research contractors). Organizations such as the Electronic Frontier Foundation (http://www.eff.org) were created to defend free speech and openness on the Net and to provide resources for individuals who wished to learn more about privacy, copyright, and intellectual property issues in the new electronic environment[3].

If the 1970s was the decade in which the computer-science research community discovered and built upon the ARPANET, the 1980s was the time when this experience began to be generalized to other branches of science. Nobel laureate Joshua Lederberg had pointed to this potential as early as 1978[4], and he was later instrumental in promoting the notion of network-based "collaboratories"[5]—a concept that has begun to take off in scientific communities[6], including medical informatics[7].

By the middle of the decade, the generalization of the technology and its increasing maturity led to the gradual transfer of its oversight from the Department of Defense to the National Science Foundation, where it was known for a time as NSFnet. Parallel networking activities, such as CSnet for the non-ARPA-related computer-science community and BITNET for other academic institutions, eventually merged and the resulting conglomeration of networks adopted the Internet name.

Acceptance of the role of the Internet in science failed to lead to much interest in the technology within the health-care-delivery community. Practitioners largely remained unaware of the Internet, and the only health centers that tended to be connected to the national network were those affiliated with research universities, in which case their network connection was typically "borrowed" from their main campus.

Several federal agencies played major roles in the evolution of the Internet and the development of policies regarding its use in the late 1980s and early 1990s. Most prominent among these were the Department of Defense (DOD, and in particular its research arm, ARPA*), the Department of Energy (DOE), the National Science Foundation (NSF), and the National Aeronautics and Space Administration (NASA). These agencies, and others with networking interests, formed the Federal Networking Council (FNC), which in turn had an advisory group from the private sector, known as the FNC Advisory Committee (FNCAC). The Department of Health and Human Services has been represented on the FNC by the National Library of Medicine (NLM).

* Over the years, the name of ARPA has switched between DARPA and ARPA several times, depending on whether the administration has wanted the word "Defense" associated with the agency's title.

GROWTH IN THE 1990s

It was in 1989 that Tennessee Senator Al Gore began to promote the notion of a new national research program that would promote the technology of the Internet and bring it to a level of quality and sophistication that would attract an even larger segment of society. He argued that such technologies could address major societal needs while promoting US economic competitiveness. He gradually built bipartisan support for legislation in the area, which was eventually signed into law as the High Performance Computing and Communications (HPCC) Act of 1991. When Gore's father had been a senator in the 1950s, he was instrumental in passing legislation that led to the creation of the interstate highway system in the United States. This analogy led some observers to dub the Internet Gore's "information superhighway", a nickname that became heavily used in the first part of this decade.

The political process required to gain support for the HPCC initiative from Congress required a substantial educational effort. One tool in this effort, continued to the present, has been the development of an annual "blue book" which outlines several societal "grand challenges" and argues for the role of high performance computing and communications in achieving those goals. Many of the examples from these books have been drawn from biomedical science. The annual reports are now placed on the Web for public review as well as distributed in printed form (see <http://www.ccic.gov/about/>).

With the passage of the HPCC legislation and the initiation of the resulting research program, which included all the agencies involved with national networking and computing, it was necessary to create an office that would help to coordinate the cross-agency activities. The first director of the National Coordinating Office (NCO) for HPCC was Dr. Donald A.B. Lindberg. Already playing a key role in the medical community as Director of the National Library of Medicine, Dr. Lindberg agreed to take on the additional responsibilities associated with the NCO directorship, and he established its first office on the grounds of the NLM in Bethesda, MD. Having the NCO for HPCC at the NLM helped to make clear the link between medicine and the new research programs, and some of the research dollars were appropriated for advanced networking programs and testbeds that were promoted by the NLM[8]. In time, however, it became clear that the NCO directorship was a major responsibility and that it required a full-time commitment. With Dr. Lindberg's resignation and the appointment of new leadership, the NCO moved to the offices of the NSF in Arlington, VA.

The responsibilities of the NCO are much broader than networking alone. Coordination among agencies is promoted not only in Large Scale Networking (LSN) but also in High-End Computing and Communications (HECC); High Confidence Systems (HCS); Human Centered Systems (HuCS); and Education, Training, and Human Resources (ETHR). Coordinated working groups in all these areas involve agency representatives from several of the participating government organizations. The NCO provides support to the Committee on Computing, Information, and Communications (CCIC), the overall oversight group for such topics within the Office of Science and Technology of the President (OSTP). Information about the NCO, its mission, and its relationship to other government organizations can be found on the web at <http://www.ccic.gov/>.

Of course the greatest change in the Internet environment of the 1990s has been the introduction and rapid adoption of the World Wide Web. This phenomenon has had a remarkable impact on our global society in just a few short years[1, 9]. The penetration into our homes, schools, and workplaces has arguably exceeded the rate of adoption of earlier popular consumer technologies such as television and video-cassette recorders.

By April 1995 the Internet had been fully "privatized" and was no longer dependent on federal funding for any component of the backbone. It is an especially interesting and impressive example of the way in which, over time, a speculative government research program can lead to technologies and systems that are commercially viable and no longer require government support. The increasing use of national networking by society, stimulated not only by the Web but by the eventual decision to welcome commercial hosts to the Internet community, has resulted in projections that Internet traffic on commercial communications systems will soon exceed the traffic derived from traditional voice telephony*. With explosive growth in other communications technologies, ranging from high-speed modems and cable modems to wireless communication systems and satellites, it has become clear that the communications vendors of the future will deal with products and services we have only begun to contemplate. The Telecommunications Act of 1996 was intended, in part, to deregulate the industry so that novel alliances and new methods of communication could be more effectively introduced into the society.

INVOLVEMENT OF THE MEDICAL COMPUTING COMMUNITY

Given the 30-year story of the Internet just summarized, it is natural to ask just how and when the medical informatics community became involved. In the early 1970s, when the ARPANET was still young, two medical computing groups were affiliated with computer science departments that were among the earliest users of the network. At Stanford University there was an active existing collaboration between artificial intelligence (AI) researchers from the Computer Science Department (notably EA Feigenbaum and BG Buchanan) and from the Departments of Chemistry (C. Djerassi) and Genetics (J. Lederberg). Working first on a system to infer organic structures from mass spectral data (the Dendral program[10]), and later on clinical problems in diagnosis and therapy planning[11], they proposed the creation of a computing resource that would be shared among a national community of researchers interested in AI applications in biomedicine. The resource, known as the Stanford University Medical Experimental computer for AI in Medicine (SUMEX-AIM) was funded in 1973 by the Division of Research Resources (DRR) at the National Institutes of Health (NIH). With the help of the DRR, the SUMEX machine became the first non-DOD-funded machine connected to the ARPANET. This resource continued for almost 20 years and supported a wide variety of collaborative research activities that depended upon the ARPANET for access, including Internist/QMR[12, 13].

A sister AI-in-Medicine machine was funded by the DRR a few years later at Rutgers University. The Rutgers Resource similarly supported collaborative research and featured a connection to the ARPANET.

At both Stanford and Rutgers, much of the network use was focused on remote logins, since the computers themselves were being made available to distant users who did not have similar resources on their own campuses. However, email rapidly became

* Personal communication, Dr. Vint Cerf, MCI Communications, Inc.

a major element in the community building that occurred, leading to Professor Lederberg's prescient observations about the role of the network in support of scientific research activities[4].

By the late 1970s, other university-based biomedical computing resources were beginning to join the ARPANET club, but it was not until the 1980s that the greater biomedical community began to use the national network. The NIH (with the exception of the NLM) was slow to realize the importance of connecting to the Internet and came online much later than did most of the academic research institutions that it funded and with which its scientists were interacting.

In 1986, shortly after Donald Lindberg assumed the directorship of the NLM, several planning panels were commissioned to help to develop a 10-year plan for the Library. One panel proposed the role of electronic information in support of biomedical sciences[14], and it was this insight that led in time to the creation of the NLM's National Center for Biotechnology Information (NCBI) and, arguably, to the emergence of bioinformatics as a discipline.

A second panel was charged with providing advice in the field of medical informatics, and among their recommendations was one that dealt specifically with electronic communications[15]. They noted that "only small segments of the biomedical research community have access to the integrated computing and network communications services that are essential to future medical information systems." They accordingly urged the NLM to work to assure that "by the end of the next decade, there will be a national computer network for use by the entire biomedical community, both clinical and research professionals. The network will have advanced electronic-mail features, as well as capabilities for large file transfer, remote computer log-in, and transmitted graphics protocols. It will either be part of a larger national network of scientists or will have gateways to other federally sponsored networks" (p. 65). By 1996 we did indeed have the World Wide Web and much of what the committee had proposed.

However, in 1989, frustrated by the slow movement of the biomedical community in areas related to wide-area networking, I personally began to promote the notion that we needed more effective federal leadership from Health and Human Services (HHS), especially in the area of the role of the Internet in health-care delivery. This was the theme of a talk that I gave in the opening session at the Symposium on Computer Applications in Medical Care (SCAMC) in 1989 as well as a similar address that I presented in 1990 at the annual meeting of the Society for Medical Decision Making[16]. My concern was that the health-care community, functioning as it does in multiple, separate organizations without central coherence, has a special need for federal guidance in understanding and suitably adopting this kind of complex technology. Yet there was no similar emphasis on networking policy and involvement in HHS that we were already seeing in other "mission-oriented" agencies such as NASA and DOE. I felt it was clear that wide-area networking was just as important to the present and future of health care as it was to energy management and space exploration.

There followed, in the 1990s, rapid adoption of Internet technologies in the biomedical research community, especially after the introduction of the World Wide Web. In addition, the public's thirst for health information has been demonstrated by their aggressive use of the Web in exploring medically-related sites. Federal health agencies, like the rest of government, have moved to develop online resources for essentially every agency (with major efforts by NLM, NIH, and the AHCPR). The NLM has offered a connections grants program to encourage hospitals to link to the Internet, but there otherwise is no coordinated federal effort to bring together health-care organizations in areas related to the Internet and its potential clinical use.

RECENT DEVELOPMENTS

With the end of the HPCC's initial five-year life span, the Clinton administration has sought to define what the suitable next phase should be in the evolution of the federal R&D program in this area. Many of the President's speeches have pointed to the role of the Internet in education, for example, where he has expressed a strong commitment to wiring the nation's schools in the coming decade. On the research side, he has promoted a new research program that has been dubbed the Next Generation Internet (NGI). The President's budget for FY98 included $100M in incremental funding for NGI-related research, with those dollars distributed principally to the four key agencies mentioned earlier. The final funding was somewhat less than that, largely due to removal of the proposed funding for the Department of Energy, but the NLM did get brought in with a $5M component. The new FY99 budget also includes $105 for the NGI, of which $5M is earmarked currently for the NLM. There is evidence that the program has strong bipartisan support in both houses of Congress.

There has been some confusion about the nature of the NGI program because some have seen it as simply the creation of a newer, faster Internet and they have wanted to be sure that they (or their constituents) are included in any connections program. The approval of the program was delayed in 1997 partly because of concerns that the NGI would create a nation of "haves" and "have nots" where rural areas, or universities other than the major research centers, would be left behind. The NCO and federal agencies created an implementation plan to clarify the research goals of the program as well as the plans for spending the appropriated funds (see <http://www.ccic.gov/ngi/implementation/>).

As the NGI program was being proposed by the federal government, a consortium of research universities was forming to address issues of Internet support for academic research. Members of the consortium agreed to make major upgrades to their campus networks and then proposed to work together to assure high-bandwidth connectivity among their campuses. As the "regular Internet" has become congested with routine, non-scientific use, there has been a growing sense of the need for a more protected, or higher quality network that could support research (as the original ARPANET did in the past). The original consortium called themselves Internet-2, which led to confusion in Congress about the relationship between Internet-2 and NGI. With the involvement of well over 100 universities, the consortium recently incorporated and is now formally known as the University Consortium for Advanced Internet Development (UCAID; see <http://www.ucaid.org>).

The "alternate network" to which the Internet-2 organization initially sought connectivity was an NSF-funded network, overseen by MCI Communications and known as the vBNS (Very high Bandwidth Networking System), which had been created to connect the NSF-funded supercomputers in Illinois and California. Recently, however, UCAID has broadened their infrastructure options to include a new network called Abilene being developed jointly with commercial partners (Cisco Systems and Qwest Communications). Details are available at the UCAID web site.

Part of the original HPCC legislation, but implemented only in early 1997, was the creation of a special Presidential advisory committee to assist the White House and OSTP with planning and policy in the area of national information technology research programs. The Presidential Advisory Committee on High Performance Computing and Communications, Information Technology, and the Next Generation Internet (they are still hoping for a shorter, more pronounceable acronym than PACHPCCITNGI!), has members from the private sector, generally from industry or academia. The Committee provided expert witnesses for Congressional hearings on the NGI legislation in 1997 (see September 10, 1997 hearing of the full committee at <http://www.house.gov/science/hearing.htm>) and has prepared a major report for the President which is due to be released shortly after this paper goes to press. The details of advisory committee activities can be found on the NCO's web site at <http://www.ccic.gov/>.

RESEARCH AGENDA

The Presidential Advisory Committee has identified several major areas in which research is needed relative to the future of the Internet and high performance computing. This agenda is consistent with the categories and emphases identified by the federal agencies in their implementation plan:

- Methods for scaling the Internet to meet the needs of a global society in which essentially everyone is connected, both at home and at work. Issues include improvements in quality of service (QOS) with decreased latency, increased bandwidth with service guarantees, new approaches to pricing, and methods for addressing the issue of the "last mile" (the connections between our homes/offices and the Internet, whether they be via physical, modem connection or via wireless methods including satellites).
- Creating the applications that will drive our understanding of what technical challenges remain. Health care provides many suitable challenges, for example, as do commerce and education. Note, however, that the value of these applications to the total effort will lie in specific efforts to extract generic lessons from the work rather than to focus on the application's domain performance alone.
- Inventing and building the "devices" that will provide our connectivity to the networked society, ranging from the future of personal computers to personal digital assistants and environments in which microprocessors are

ubiquitous in objects from our environments (such as elements in "smart houses" or similar environments accessible via the Internet).

- New generations of software, and especially middleware, which are not keeping up with the rapid changes in the underlying technology and which many believe are not adequately appreciated as an area in profound need of research investment.
- Research in high-end computing and computation that will work in tandem with the national network. The near-demise of the US supercomputer industry raises troubling questions for policy as well as research. Should there be a US supercomputing industry? What is the risk to the country if we let the technology wither, or let other nations rise to dominance in the area? If it is concluded that the limited demand for supercomputing prevents such technology from being commercially viable, is it in the national interest, for the foreseeable future, for the government to subsidize and promote the activity?
- What are the suitable economic models for the future networked society and how should resulting insights affect federal regulatory philosophy and approaches? We need to understand better how to divide networking activities realistically between the public and private sectors, maintaining incentives for commercial investment while addressing issues of access and pricing that will allow all citizens to participate in the benefits of the new technologies.
- And then there are social and ethical concerns that warrant study. How will our society benefit from the transformation with which we are involved? What are the risks? Can we document changes that are either positive or worrisome? Can we define policies, procedures, and technologies that will help us to avoid the pitfalls? There are many important health and health-care examples, of which data privacy and confidentiality is a particularly prominent example[17].

WHAT LIES AHEAD

Given the bipartisan support for the NGI program in Congress, it seems likely that federal research investment in the future of the Internet will continue in the coming decade. The research program will undoubtedly be accompanied by Congressional efforts to assure that traditionally underserved regions and schools are not left out as the Internet advances and improves.

The commercial sector will continue to invest heavily in the Internet, both as users of the technology and, for the telecommunications companies, as service providers and innovators. The rapid rise of the Web has shown us, however, that it would be folly to try to anticipate the rate of change or the new technologies that may arise in the decade ahead. We should probably look to industry for incremental change, and for efforts to make the technology more robust, while academia and science will continue to be the source for paradigm shifts (such as the Web) that will subsequently be taken up by the commercial world. Interactions with regulatory policy will be extremely important, as in, for example, the resolution of "last mile" connections to homes in affordable ways.

But what of research? What will be filling the pipeline for 20-30 years hence in the way that the networking investment by ARPA did in the 1960s and 1970s? The nation must have a balanced research portfolio in information technology, supporting both short-term demonstrations and longer term innovation and technology development. We are in an era when Congress has been much more focused on short-term benefits from research investment, and many observers believe that even the Internet is ample evidence of how shortsighted that view of research can be. Medical informatics research investment must be similarly balanced between basic and more applied investigations. We will be lost if we demand short-term payoffs from all research activities.

The health-care community could be doing much more with the networking environment that we have in place today, but must recognize (a) the forces that are preventing optimal cooperation among our organizations given an inherently distributed, competitive environment; (b) the logistical barriers to systems integration, largely in the area of standards development for both data exchange and terminology; and (c) the difficulty in justifying institutional investment by demonstrating cost-effectiveness in an environment where intuition is not enough but formal experiments are often flawed or impossible to perform.

Despite these problems, the future of wide-area networking for the health-care community is exciting[18]. We in medical informatics have every reason to support the NGI effort and to contribute to it aggressively.

CONCLUSIONS

Those in the informatics community must become visible, credible players in health-care planning and policy, demonstrating to our institutions the strategic role that technology will play in the future of health care and seeking positions of influence so that the informatics perspective is well represented in high-level decision making. We must be much more involved in the political process than we have been in the past, and recent efforts by AMIA's Public Policy Committee are encouraging evidence that the profession is starting to play this kind of role.

Similarly, HHS needs to take more seriously the need for its aggressive involvement in the area of national networking, rectifying its current lack of involvement in the federal networking program and in its development. The NLM cannot be expected to handle these issues for all of HHS. Informatics researchers need more dollars for investigation, of course, but we also need to help effect a cultural change in the health-care community, especially among leaders, including those in HHS. They, with our help, must persuade the core science agencies (as NASA and DOE have done) that health-care applications can help to drive the underlying science of computing and communications. To make the argument, we must continue to do the kind of work that will show the national information-technology research community that there are important generic contributions from both our basic and applied research efforts.

References

1. The year of the Internet. Time 1995 December 25:21-46.
2. Kahn R. The role of government in the evolution of the Internet. Communications of the ACM 1994;37(8):15-19.
3. Gelman R, McCandlish S. Protecting Yourself Online: The Definitive Resource on Safety, Freedom & Privacy in Cyberspace. Harper Edge, 1998.
4. Lederberg J. Digital communications and the conduct of science: The new literacy. Proceedings of the IEEE 1978;66(1):1314-1319.
5. Cerf V, and the Committee on a National Collaboratory, Computer Science and Telecommunications Board (National Research Council). National Collaboratories: Applying Information Technology for Scientific Research. Washington, D.C.: National Academy Press, 1993.
6. Kouzes R, Myers J, Wulf W. Collaboratories: Doing science on the Internet. IEEE Computer 1996(August):40-46.
7. Shortliffe E, Patel V, Cimino J, Barnett G, Greenes R. A study of collaboration among medical informatics research laboratories. AI Med 1998;12:97-123.
8. Lindberg D, Humphreys B. The High-Performance Computing and Communications Program, the national information infrastructure, and health care. JAMIA 1995;2:156-159.
9. The Internet. The Economist 1995 July 1:S1-S18.
10. Buchanan B, Feigenbaum E. DENDRAL and Meta-DENDRAL: Their applications dimension. Artif Intel 1978;11:5-24.
11. Buchanan BG, Shortliffe EH, eds. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Reading, MA: Addison-Wesley, 1984.
12. Miller RA, Pople HE, Myers JD. INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine. New Eng. J. Med. 1982;307:468-476.
13. Miller RA, McNeil MA, Challinor SM, Maserie FW, Myers JD. The Internist/Quick Medical Reference project: Status report. West J Med 1986;145:816-822.
14. Planning Panel Number 3. Long Range Plan on Obtaining Factual Information from Data Bases. Bethesda, MD: National Library of Medicine, 1986:
15. Planning Panel Number 4. Long Range Plan on Medical Informatics. Bethesda, MD: National Library of Medicine, 1986:
16. Shortliffe EH. Medical informatics and clinical decision making: The science and the pragmatics. Med Dec Mak 1991;11(4):S2-S14.
17. National Research Council. For The Record: Protecting Electronic Health Information. Washington, D.C.: National Academy Press, 1997.
18. Shortliffe E. Health care and the Next Generation Internet (editorial). Anns Int Med 1998;129(July 15).