Leadership

By George Strawn

You Ain't Seen Nothin' Yet

Information technology has been changing the academy for the past fifty years, but those changes are small compared with what we can expect in the next fifty years. Three relatively recent IT "subrevolutions" have had an increasing impact on research, scholarly communication, and education.

The *PC subrevolution* began in the 1980s. This subrevolution was made possible by the mid-1950s invention of the solid-state circuit, which led to the now-ubiquitous computer chip. Thus it was that the room-size computer of 1970, costing several million dollars and processing one million instructions per second, became the lap-size computer of 2000, costing several thousand dollars and processing one billion instructions per second. The increase in computer price-performance and the decrease in size did not, of course, translate into a cost decrease for higher education. A 25,000-student institution that might have had several computers in 1970 was well on the way to having more than 25,000 computers by the end of the century.

The *Internet subrevolution* began in the 1990s, although its prehistory dates back to 1969, when the first three nodes of the Department of Defense's ARPANet came online. In 1983, the first version of the ARPANet software was taken down and the second version, TCP/IP, was installed. In 1985, the National Science Foundation (NSF) decided to use the TCP/IP protocols for NSFNet, which it was building to connect higher education to its new supercomputing centers. By 1995, when NSFNet was retired in favor of commercial Internet providers, the backbone network bandwidth had grown from 56 bits per second (bps) to 45 million bps. In the ten years since, backbone bandwidths have reached 10 billion bps. Network traffic has also grown exponentially over this time, and the number of Internet users is approaching one billion. If the effect of the PC subrevolution was to turn hundreds of thousands of large computers into hundreds of millions of small ones, the effect of the Internet subrevolution was to turn those hundreds of millions of PCs into one giant global multiprocessor computer. This subrevolution also turned what was originally a computing device into a communicating device.

IT-assisted education is currently an early-phase disruptive technology.

The *information subrevolution* began (according to my account) in the 2000s and is based on the PC and Internet subrevolutions. Now that computers and networks are sufficiently ubiquitous, increasingly it is the information that they contain and transmit that is the focus of new developments. World Wide Web protocols have enabled people to create and access an ever-increasing hoard of Web pages containing information on just about everything. The Web and Google make information more accessible, more open, and more free. Open information for computers, called *open source,* has become a force in the software world. A newer, similar force is service-oriented architecture (SOA), which can be implemented by a set of protocols called *Web services*. SOA aims to make software modules available for reuse over the Web in a highly automated manner. Open information for humans is also exploding, in part based on a read *and* write Web technology called a *wiki*. *Wikipedia*, a prime example, is an open encyclopedia developed by countless contributors using the open-source MediaWiki software. Also, Web logs (known as *blogs*) are written by millions of people and read by tens of millions.

Effect on Research

The original computer revolution began in the 1950s and produced what we used to call *mainframe computers*. From their inception, they made an impact on academic research, especially in the physical sciences and engineering. Numerical solutions to an increasing array of mathematical equations became computable, and computational science and engineering subsequently became important specialties in many disciplines. When the PC subrevolution occurred in the 1980s, colleges and universities curtailed purchases of mainframes and began buying PCs (and their more powerful cousins, scientific workstations). The "Lax Report," written in the early 1980s, called attention to this fact and to the consequent loss of access by researchers who needed large-scale computers.[1]

The Lax Report helped influence Congress to counter this trend by directing the NSF to create supercomputing centers for academic researchers. As mentioned above, the NSF not only created supercomputing centers but also connected them to university and college campuses via the concurrently created NSFNet. As PC chip technology improved, the architecture of supercomputers changed greatly. Today, many of the largest computers used for science and engineering research are assembled from hundreds or thousands of PC-like chips. We are only a few years away from "petascale" computing systems capable of 10^{15} operations per second (one peta-op), which will be achieved by approximately one hundred thousand processors each capable of ten billion operations per second. In this ultra-high-performance environment, some researchers speak of computational science and engineering as "the third mode" of science, along with the experimental and theoretical modes. Others prefer to say that such systems permit a novel combination of experimental and theoretical modes and that much new theory is expressed in software rather than in equations. Perhaps the premier application of supercomputers is the increasingly accurate modeling and simulation of complex systems such as cells and organisms, societies and global climate, stars and galaxies. The NSF has recently recognized the increasing importance of information technology to the science and engineering research enterprise by elevating its Cyberinfrastructure Division to the Cyberinfrastructure Office under direct management of the director of the NSF.

Effect on Scholarly Communication

In 1964, John Kemeny, then chair of the Dartmouth mathematics department, published a lecture that he had presented at MIT in 1961: "A Library for 2000 A.D."[2] Although he was looking forward with only very early IT equipment to guide him, his vision—of a universal repository for scholarly literature that could be accessed and searched from remote terminals around the world—is now coming to pass. An early step was taken in 1991 when Paul Ginsparg's ArXiv (http://arxiv.org/) began serving the physics community with an electronic preprint service that created a new, simplified way to access scholarly literature. Today, as budgets have tightened, some research libraries have more electronic journal subscriptions than paper ones. Currently, at least two business models are being tested to determine the best way to support more open access to scholarly literature: (1) the author pays/free access model; and (2) the free access after *n* months model. Also, the National Institutes of Health (NIH) has asked principal investigators (PIs) to submit copies of published articles that have resulted from NIH-supported projects to an NIH-maintained open-access "active archive."

Open access to the world's scholarly literature is just the beginning. Now that the digital library is being realized, text processing is increasingly being applied to tease out meaning ("machine understanding") of that literature. Because of the explosion in the amount of information available, getting computers to "understand the information" so that humans can be more precisely directed to what they are looking for, or so that humans can be taken out of the loop entirely, is essential. There is just too much information for human attention. The biomedical area is taking the lead in this activity (for example, see *Ontologies for Bioinformatics* by Kenneth Baclawski and Tianhua Niu). With suitable text processing of biomedical articles, computer systems are now able to make their own e-scientific discoveries: for example, the pharmaceutical industry is beginning to use such techniques for drug development. As *metadata* (annotation data about information) is increasingly created for digital information (and this will accelerate when techniques for automating the creation of such metadata are perfected), and *ontologies* (machine-readable delineation of the concepts and relations in a subject area) are developed for scientific fields, precision searches will be possible, and text analysis will be increasingly used to pursue scientific discoveries.

Effect on Education

Computationally based research has existed for more than half a century, and digital libraries are currently under development. On the other hand, IT-*based* (not just IT-assisted) education will be a future development. This is not to say that current educational support such as instructional management systems, Google, and wirelessly networked laptops don't contribute to the educational environment. But to use the phrase adopted by the NSF for one of its recent initiatives, we are just beginning to create "a science of learning."[3]

IT-assisted education is currently an early-phase *disruptive technology*, as defined by Clayton Christensen.[4] Disruptive technologies start out by providing a service (or product) that is *poorer* than the ones currently used. (And surely faculty members today are justified in thinking that traditional educational methods, at their best, are more effective than IT-assisted methods such as distance education.) Christensen has studied disruptive technologies that improve at a more rapid rate than the incumbent services/products, and incumbent providers are often caught off-guard because they have ignored the poorer disruptive technology. By the time they turn serious attention to the disruptive technology, other providers have already established viable business plans and processes to market and support the new, now-better-than-current-services disruptive technology. In the case of higher education, Christensen notes that for-profit universities, community colleges, and corporation training programs may all be more receptive to IT-based learning because of their focus and emphasis on cost containment.

* * *

The IT subrevolutions discussed above have already affected higher education and will certainly affect it more in the future. If the changes discussed here sound extreme, they should be compared with those predicted by Ray Kurzweil, the noted IT inventor and futurist. In his 2005 book *The Singularity Is Near,* he makes the case for IT-induced changes that could occur throughout the world over the next fifty years and that would truly lead one to say, "You ain't seen nothin' yet."

Notes

1. Peter D. Lax, ed., "Report of the Panel on Large-Scale Computing in Science and Engineering," Coordinating Committee NSF/DOD, December 26, 1982 (the Lax Report), <http://www.pnl.gov/scales/docs/lax_report_1982.pdf>.
2. John G. Kemeny, *Random Essays on Mathematics, Education, and Computers* (Englewood Cliffs, N.J.: Prentice-Hall, 1964).
3. See <http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5567&from=fund>.
4. See Clayton M. Christensen, Scott D. Anthony, and Erik A. Roth, *Seeing What's Next: Using the Theories of Innovation to Predict Industry Change* (Boston: Harvard Business School Press, 2004).

George Strawn is CIO at the National Science Foundation.