POLICY FORUM: COMMUNICATION

Scientific Collaborations at a Distance

Stephanie Teasley* and Steven Wolinsky

The Internet has the potential to enhance collaboration among researchers by facilitating rapid dispersal of information and the coordination of numerous, complex, real-time interactions. The most frequently used applications of the Internet support asynchronous transfer of static text files and images from large publicly available databases. However, single-user access to static information is a small component of the possible spectrum of activity.

A growing number of telemedicine and telemicroscopy projects allow researchers to control experimental equipment remotely in real-time. For telemicroscopy, the instrument operators or local researchers give control of the instrument to the remote investigator. Investigators examine data they have generated themselves, rather than accessing community-held data resources. Even though the process is more interactive than static file transfer, applications for remote equipment access are not totally interactive, as the flow of information is typically unidirectional.

The Internet also has the power to support scientific collaboration by linking investigators who can then interact with each other in real-time while dynamically manipulating data. Coordination of most geographically distributed research collaborations now is based on traditional means of communication—the telephone, facsimile transmissions, and in-person meetings. However, the time delays inherent in asynchronous means of communication hinder progress and can create opportunities for misunderstandings or extended negotiations between collaborators who need to work tightly together. Geographically distributed collaborators are often faced with heavy travel demands and/or endless audio conferences. Recent experiences of scientists using collaboration technology illustrate how this approach can support effective collaboration at a distance.

Collaboration at a Distance

An important paradigm of Internet-mediated science is the collaboratory, or "laboratory without walls," a virtual entity created by means of a computer network (*1–5*). Researchers in physical oceanography (*6*), worm genomics (*7*), and space physics (*8*) were the first to recognize the potential of a collaboratory and to implement the earliest prototypes (*9*). Over time, advances in networking technologies have improved the data-carrying capacity of the network, the transmission time, and the level of availability of the Internet to the world at large. Such technical improvements in computer hardware, software, and infrastructure have increasingly optimized conditions for the application of Internet-based tools to support collaborative research efforts. Indeed, the necessary networking components for many aspects of a collaboratory are commercially available today and do not require high-end hardware or high-level technical support. Software costs can be quite modest as many applications now bundle "group productivity tools" into standard packages. Costs increase as specialized needs of the project are identified, and as the technical expertise needed for installing and supporting these specialized applications grows.

> "... distributed work will ultimately require a fundamental transformation of scientific practice ..."

One example of a working collaboratory is the Great Lakes Regional Center for AIDS Research (CFAR). The Great Lakes Regional CFAR is a consortium of scientists at Northwestern University, the University of Minnesota, the University of Michigan and the University of Wisconsin–Madison who have consolidated their complementary scientific expertise. Core facilities allow investigators to join proteomics, genomics, bioinformatics, animal models, and clinical studies into a unified whole.

This collaboratory works under the organizing principle that technical infrastructure follows function. A needs assessment was undertaken by social scientists at the beginning to assess the diversity of requirements and the extent of commonality in practice. Among the priority requirements were real-time audiovisual communications with document sharing, remote control of experimental equipment, information search and retrieval, and transfer and storage of images and large data sets. After evaluation of the scientific practice and available computer hardware, commercial applications were selected. Actual collaboratory costs were modest, as the required functionality did not involve any customized software development. Technical end-user support at each site has ensured optimal collaboratory use. The technical staff installed and configured the software, and conducted site-to-site pilot tests to resolve any problems. They then trained the scientists and sat in on the first uses of the tools to help troubleshoot any problems.

A Web site (www.greatlakescfar.org) functions as a network resource to catalyze and coordinate interactions. At the Web site, participants are informed about core services offered, survey data are collected, new members are registered, past presentations are archived, observations are recorded, help documents and announcements of upcoming events are provided, and new tools are demonstrated.

Multiparty collaboratory-supported functions use a number of applications, for example, video conferencing combined with real-time document and image sharing, to provide the participants in a laboratory meeting with the tools to see each other and to view shared text and images on their personal computers (see the figure on page 2255). Here, a tissue section imaged by a shared microscope at one site was broadcast to another site where the biopsy was performed. A participant at the distant site examines the tissue section by manipulating the microscope stage through a remote-control application. A pathologist interprets histological findings, and everyone discusses the results and implications in real-time.

The collaboratory also supports distance learning by delivering real-time broadcasts of lectures through a virtual presentation application combined with a teleconference. Participants with an Internet connection log in from any location through the Web site portal. Teleconference is necessary because there is no reliable, high-quality Internet application for multiparty voice transmission.

Files are conveniently exchanged across sites, and data from joint research projects are stored in a centralized location. Although no computer is completely secure against intrusion and malicious attack, encryption and security protocols provide user authentication to protect patient files and to maintain

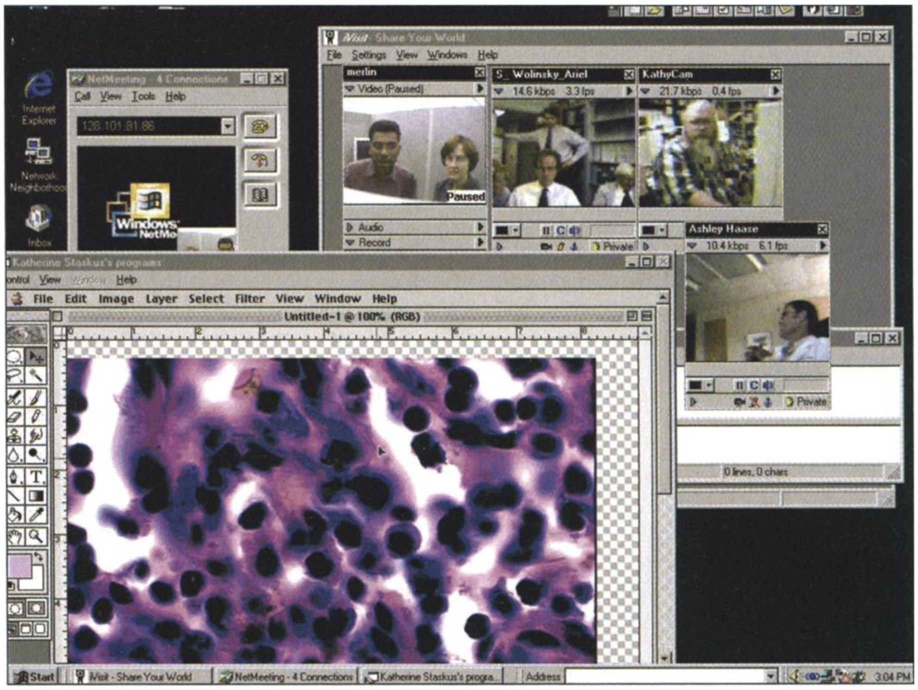
S. Teasley is with the Collaboratory for Research on Electronic Work, School of Information, University of Michigan, Ann Arbor, MI 48109, USA. S. Wolinsky is in the Department of Medicine, Northwestern University Medical School, Chicago, IL 60611, USA.

*To whom correspondence should be addressed. E-mail: steasley@umich.edu

secure communication across the Internet.

A collaboratory improves coordination and collaboration among scientists by enhancing communications and facilitating access to information, but it has its limits. Social and practical acceptability are the primary challenges of a distributed problem-solving environment. A collaboratory cannot replace the richness or the commitment engendered by face-to-face interactions. As in other collaborative arrangements, concerns surrounding trust, motivation, and normative practice for data access, ownership, and acknowledgment can hinder collaboratory function (*10*).

Principal technical challenges of collaboration via the Internet are its restricted data-carrying capacity and the time needed to deliver distributed high–bit rate audio and video between locations on the network. Although the Internet has sufficient capacity, there is no assurance that the high bandwidth and rapid response time required for sustained performance will be consistently available when needed as other traffic competes for the same capacity at the same time. Like freeway traffic, the Internet is subject to high-use periods when the sheer volume of users slows performance.



Participants in the Great Lakes Regional CFAR review results by videoconference with real-time sharing of documents and images of histological lymphoid tissue sections from a remotely controlled microscope. As a distant participant manipulates the microscope stage, a pathologist interprets histological findings, and everyone discusses the results and implications in real-time.

Other limitations of the Internet are those inherent to the available applications, especially when used across diverse platforms. In addition, many of the available collaboration tools are still somewhat awkward to use, primarily because designing software effectively for group use is a new area.

Opportunities and Challenges

The information technology challenge is how to take advantage of the high-performance Internet capabilities now available from the complementary University-led Internet2 and the federally run Next Generation Internet (NGI) initiatives. Internet2 is a consortium of more than 180 American Universities and 70 corporations working in cooperation with NGI to develop and deploy advanced network applications and new technologies (www.internet2.org). These initiatives provide access to advanced backbone networks, such as Abilene, the very high performance Backbone Network Service (vBNS), and the Energy Science Network (ESNET), to ensure a cohesive high-performance network infrastructure among government, industry, and academia.

Although advanced networking initiatives promise to provide the high-speed communication and the quality of service required by advanced applications, end-to-end connectivity is "the last-mile problem." Investment of financial resources and intellectual capital needed to establish on-campus connectivity for every building and laboratory is more than many universities and some government facilities can afford. Wider use by the academic community will require the influx of federal funds from those agencies participating in the advanced networking initiatives and a concomitant drop in price by the network service providers.

Many U.S. federal agencies that support academic research recognize the challenges and opportunities and, like the NIH's National Center for Research Resources, have created collaboratory supplements for biomedical projects. A recent report from the European Technology Assessment Network (ETAN) also calls for increased development of collaboratories, high-performance networks, and digital libraries (*11*).

This approach to distributed research has advantages and disadvantages. Collaboratories can greatly increase the volume and quality of data available to collaborators. But they increase the need for new ways to keep track of shared data and complicate issues of data ownership. Collaboratories also create opportunities for more interaction with distant collaborators without the need to travel, but this requires coordinatation among the participants. The nature of the work to be accomplished, the scientists' readiness to collaborate, and the personal and institutional level of technology support are all factors that influence collaboratory deployment.

Successful distributed work will ultimately require a fundamental transformation of scientific practice at the social level. Who has access, who can participate, who benefits, and how to reward truly shared scientific achievement are outstanding questions for those using collaboratory tools. Credit for receipt of competitively awarded grants from federal agencies, the equitable division of indirect costs across collaborating institutions, and the responsibilities that the organization assumes in a distributed problem-solving environment are questions that universities will eventually need to answer.

References

1. T. A. Finholt, G. M. Olson, *Psychol. Sci.* **8**, 28 (1997).
2. J. Lederberg, K. Uncapher, *Towards a National Collaboratory: Report of an Invitational Workshop at the Rockefeller University* (Directorate for Computer and Information Science, National Science Foundation, Washington, DC, 1989).
3. National Research Council, *National Collaboratories: Applying Information Technology for Scientific Research* (National Academy Press, Washington, DC, 1993).
4. W. A. Wulf, *Science* **261**, 854 (1993).
5. A. M. Weinberg, *Science* **134**, 161 (1961).
6. B. W. Hesse, S. Sproull, S. B. Keisler, J. P. Walsh, *Commun. ACM* **36**, 90 (1993).
7. B. Schatz, *J. Manage. Inf. Sys.* **8**, 87 (1991).
8. T. A. Finholt, G. M. Olson, *Psychol. Sci.* **8**, 28 (1997).
9. T. A. Finholt, *Annu. Rev. Inf. Sci. Technol.* **36** (2001).
10. S. L. Star, K. Ruhleder, in *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (ACM Press, New York, 1994), pp. 253–264.
11. See www.cordis.lu/etan/src/topic-8.htm

12. Supported by a grant from the National Institutes of Health (P30 CA79458-03) and a gift from the James B. Pendleton Charitable Trust.