32 Collaboratories: Science over the Internet

Thomas A. Finholt

For the last decade, I have worked with colleagues at the University of Michigan and elsewhere to explore the practice of science using the Internet, focusing on a mechanism called the collaboratory. This chapter will discuss the background of this idea, some specific activity conducted at the University of Michigan, some principal projects elsewhere, and ideas about policy implications and how this concept may transform the organization and practice of science.

The term collaboratory is a combination of the words "collaborate" and "laboratory." It was coined by Bill Wulf in the late 1980s, and then popularized in 1993 by the National Research Council (NRC) report, *National Collaboratories: Applying Information Technologies for Scientific Research*. At this stage the concept is promoted principally by federal agencies, especially the National Science Foundation (NSF) and the Department of Energy (DOE). More recently, the National Institutes of Health (NIH) and the Department of Defense have sponsored collaboratory activities within some of their funded programs.

There are, of course, parallel developments in the corporate sector and some very interesting developments in the nonprofit sector. For example, the University of Michigan under the sponsorship of the W.K. Kellogg Foundation, is looking at using this technology to organize within communities in urban and rural areas.

The collaboratory concept uses media-rich technology supported over the Internet to link people with data and instrument resources, as well as with colleagues to discuss live data (see Leonard Krishtalka, Chapter 31 in this volume). These data may be the result of simulations or they

---

Thomas A. Finholt is a senior associate research scientist and director of the Collaboratory for Research on Electronic Work at the University of Michigan's School of Information. *This article is based on remarks delivered at the 26th Annual AAAS Colloquium on Science and Technology Policy, held May 3–4, 2001, in Washington, DC.*

may have been gathered from instruments. Access to collaboratories should be as transparent as possible, and should not require disruption of day-to-day activities or the mastery of complex computer applications.

At Michigan, our developmental model over the last ten years suggests that these collaboratories succeed best when accessed via Web-based interfaces. In other words, the World Wide Web has become nearly ubiquitous in scientific communities. For instance, people are very familiar with using the Web to access data resources. Any task that asks you to do something beyond that (like loading a special application, configuring a special system, or using an exotic operating environment) tends to restrict or reduce enthusiasm. Scientists are principally interested in conducting science, not in becoming computer scientists. So the Web is a nice solution that is both familiar to scientists and still supports collaborative functions.

We believe that the use of collaboratories is in an early stage now and that the penetration and application of the technology is bound to grow. There are a couple of reasons why we believe this. First, there is clearly a need for multidisciplinary activity to solve complex problems like the worldwide AIDS epidemic, identifying the Higgs boson, and so forth. These activities require integration across multidisciplinary teams with different expertise, and that expertise cannot easily be located in a single place. Solving these problems requires a mechanism that allows people to do their science with one another with ease and flexibility, traditionally achieved only through physical co-location, but now possible through the Internet.

The second factor that is going to promote the growth of collaboratories is the proliferation of tools now becoming available for people to collaborate online. For example, software developed in the private sector makes it possible to conduct meetings over the Internet, such as applications for data conferencing.

One issue that emerges as you start to think about collaboratory-based science is the potential for transformation of the research enterprise. Another is increased concern over intellectual property and ownership of content. For example, if you are carrying on an interaction mediated over the Internet, and it is catalogued in a chat stream, who owns the content of the chat stream? If you are marking up a document or a real-time image, and you come up with an interesting insight or discovery, who owns that discovery? Was it a product of your individual effort? Was it in the gestalt of the mutual activity? We will see a lot of issues of ownership and control, particularly as data resources are made available over the Web. It is likely that scientists in the future will routinely use data from collaborators they have never met in person—and that instruments at one location will be routinely used by researchers all over the world.

To better understand the collaboratory concept, I want to describe three projects at Michigan. Our past is deeply embedded in the space physics community through an application called the Space Physics and Astronomy Research collaboratory (SPARC). This enterprise was started in 1993, largely in response to the NRC collaboratory report. Initially, it was funded jointly through a cooperative agreement by the NSF Geosciences and Computer and Information Science and Engineering Directorates. SPARC's purpose was to build a system that enabled space physicists to manipulate and read data from hundreds of observatories, ground-based networks, and space-based instruments in order to study the Earth's conducting atmosphere.

Our present is represented by the first Great Lakes Regional Center for AIDS Research (CFAR). It is funded by NIH and brings together a number of institutions in the Midwest.

Our future is represented by a collaboratory being developed for the NSF's George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES). When completed in 2004, the NEES collaboratory will integrate shake tables, reaction walls, and tsunami wave tanks at 20 universities across the U.S.

The Web-based interface used in SPARC illustrates key features of a collaboratory. The top-level page has a set of summary statistics, akin to a space weather report including index values representing the current state of the solar-terrestrial system. You can also interact via a chat stream with other colleagues who might be looking at these data at the same time. And from a directory, you can branch off into hundreds of other pages created by other experts showing activity around experiments, new kinds of instrumentation and other areas.

Use of SPARC suggests three ways that general adoption of collaboratories may influence research. First, access to complementary expertise becomes much more convenient. In a typical space physics experiment, for example, a scientist may go to a remote observatory and be disconnected from his or her community and home institution. Through SPARC, a scientist can maintain that interaction. Also, people who might not have sufficient funds to actually go to the remote facility to participate in experiment can contribute their expertise via the collaboratory when it becomes most critical. For example, if phenomena change in a direction that suits the interests of a particular individual, he or she can then jump in and contribute to the ongoing research.

Second, there are implications from the point of view of students. The collaboratory becomes an excellent mechanism for supporting inquiry-based learning because students can have access to the collaboratory from the first days of their academic careers and actively engage in conducting research. In the past, they may have served an apprenticeship, analyzing archived data, dutifully producing images from some event that happened long ago, and not collecting live data until very late in their graduate careers.

A typical experiment conducted via SPARC illustrates how students benefit from using a collaboratory. In this case, a group of undergraduates at an institution in Florida participated in a live experiment using an NSF-sponsored incoherent scatter radar in Greenland. The students interacted with and were mentored by scientists located in Michigan and in California, and observed a second, independent experiment using the radar. SPARC immersed the students in an ongoing research experience and allowed them to work like real researchers, even though they were only researchers in training. Frequently, in the past, research has been off-limits to novice participants because they were not sufficiently skilled, or because resources were scarce. For instance, you cannot afford to spend scarce travel funds to send an inexperienced student to a remote observatory. However, with the collaboratory, students can participate—and gain important experience—at a relatively low cost.

Finally, collaboratories allow data from hundreds of sources to be integrated and displayed in one place. For example, in SPARC, displays from radar all over the world can be viewed by anyone with a Web browser. Through SPARC, scientists can review output from an array of instruments in real time that in the past would have been viewable only after the fact or would have required travel to dozens of separate observatories.

A recent solar event that produced aurora visible in Mexico suggests some of the possibilities of SPARC and other collaboratories. Through SPARC, users could see an animation of images taken from the POLAR satellite, which showed the aurora oval expanding and intensifying. This is a rich visual stream that students can watch in real time. Seeing this engenders great excitement—and they talk about how these images bring equations and other abstractions from the classroom to life.

To conclude this survey of the state of collaboratory development I want to describe the Great Lakes Regional Center for AIDS Research and the NEES collaboratory in greater detail. The Great Lakes Regional CFAR differs from the SPARC model in that, rather than producing custom applications, we are, as much as possible, trying to use commercial off-the-shelf technology. This project integrates a group of AIDS researchers located at Northwestern University, and the Universities of Wisconsin, Michigan, and Minnesota. This is the first distributed CFAR that NIH has funded. Participants have backgrounds in genomics, bioinformatics, biostatistics, and medicine.

The Great Lakes Regional CFAR scientists use the collaboratory in two key ways. First, software tools are used to conduct regular seminars or lab meetings among groups of principal investigators, postdocs, and graduate students across the four sites. For example, through the collaboratory, investigators at various sites can jointly analyze a tissue slide. One advantage is that the researchers can tailor the view in real time to achieve better quality control. In the past, the tissues might have been collected at Northwestern, sent to Minnesota for analysis, and a written report sent back to Northwestern. That loop does not get closed very rapidly. If you did something wrong, you would have to repeat the whole cycle to correct the error. In a real-time collaborative session, the scientists at one site can look at what is going on at another site and make immediate modifications.

A second very important and powerful activity within the Great Lakes Regional CFAR is the use of software applications to support Web-based lectures during which remote participants can make presentations to colleagues all over the world who listen and watch via their desktop computers. This has become a popular mechanism for scientists to share breaking information with one another and to present lectures. They also use this application to increase the audience for conventional lectures. For example, Bruce Walker from Harvard University recently gave a talk at the University of Minnesota. Through the Great Lakes Regional CFAR collaboratory, scientists at Minnesota were able to share Walker's talk with their colleagues at other Great Lakes Regional CFAR schools.

The University of Michigan also has NEESgrid, the collaboratory element of the George E. Brown, Jr. NEES. This system will become operational in 2004. NEESgrid is the most ambitious attempt, to date, to create an environment for Internet-mediated science.

Beyond the work at Michigan, there are other interesting efforts such as the Environmental Molecular Science Laboratory (EMSL) at Pacific Northwest National Laboratory, sponsored under the Department of Energy's DOE2000 program. The EMSL collaboratory includes televiewers to accommodate a large community of users who come in remotely to use EMSL's nuclear magnetic resonance spectrometers. Within NIH, the National Center for Research Resources is supporting a group of supplemental projects to investigate making resources Internet-accessible. A good example is the collaboratory for Microscopic Digital Anatomy at the University of California, San Diego, where you can actually steer and run their electron microscope via an Internet interface.

What are the next steps and implications? When new technology evolves, first attempts are often awkward, and it is hard to know exactly what the right form of the technology should be. That is the phase we are in with respect to Internet-mediated science. There are a lot of ideas about how it should move forward. Some ideas are focused on community-based data systems, like GenBank, while others are focused on digital libraries and electronic journals. Our vision is focused on collaboratories. We need more support for experimentation and for trials to determine what the suitable mix among these strategies might be.

With respect to further exploration, I think there is an important distinction between extrapolation and innovation. Extrapolation is simply taking what we already do and moving it into the Internet. Videoconferencing is a good example of that. But innovation happens when we are doing something that would not have been possible prior to the Internet, such as integration of hundreds of data sources into new collaborative visualizations.

To conclude, it is interesting to contrast the laboratory world we live in now with the collaboratory world we may be living with in the future. The laboratory is principally disciplinary-oriented, concerned with problems of interest within a given lab, and place-based. The collaboratory produces an environment that is more problem-oriented than disciplinary-oriented, that is global rather than local, and where research and education occurs independent of physical location.