

Running head: COLLABORATORIES

Collaboratories

Thomas A. Finholt

School of Information

University of Michigan

[chapter submitted for the *Annual Review of Information Science and Technology*, B. Cronin

(Ed.) – please do not cite without permission]

Abstract

Science is an inherently collaborative enterprise and this trend has accelerated over the past few decades. In particular, the Internet creates new possibilities for the organization of joint scientific work, specifically among geographically separated collaborators. A notable instance of Internet-mediated science is the collaboratory, or a laboratory without walls, where scientists are connected to each other, to instruments, and to data independent of time and location. This chapter explores past and current collaboratory efforts to identify factors that predict success and failure. The chapter concludes with an assessment of directions for future collaboratory development.

Introduction

Historically, joint intellectual activity has depended on physical proximity. For example, the probability of person-to-person communication under traditional circumstances is strongly constrained by distance – diminishing to zero beyond 30 meters (Allen, 1977; Kraut, Egidio, & Gallegher, 1990). In terms of a particular important kind of group intellectual activity, scientific collaboration, proximity has a direct effect on the quality and frequency of collaboration (Katz, 1994). Further, convenient access to scarce instruments provides an additional imperative for co-location in science (Hagstrom, 1965). In the high energy physics community, for instance, research activity is confined to a handful of labs worldwide (Traweek, 1992). Other scientific communities face similar limitations. The conventional response to these limitations has been residency, either permanent or temporary, at an instrument site. In terms of the performance of joint intellectual work, then, sharing a common work setting has evolved as a critical tool to support both frequent interaction among collaborators and the use of unique facilities.

Despite the benefits of co-location, there are important individual and collective costs. Individual costs involve lost productivity associated with dislocation from a familiar environment, such as when a scientist must travel to a remote facility. On a collective scale, co-location inevitably involves exclusion and isolation from those who are located elsewhere. Some mechanisms can attenuate this effect, such as conferences and workshops. But for the most part, a shared physical space comes to define a shared intellectual space where collaboration with those nearby is much more likely than with those who are distant, even after controlling for disciplinary differences (Kraut, Gallegher, & Egidio, 1990). Focusing on science, there are additional collective costs of co-location. Specifically, to the extent that co-location plays into competitive rivalries among research sites, cooperation may be undermined. For example, lack

of cooperation can lead to redundant capacity, such as several independent and under-utilized instruments, rather than a single shared and fully utilized instrument. More important, barriers to interaction across sites may slow integration of knowledge required to resolve research questions that exceed the capacity of single sites, or even single disciplines, to answer. The global AIDS epidemic, for example, is often cited as the kind of large research problem that requires unprecedented levels of cooperation from communities that have, in the past, worked independently (e.g., clinicians, bench scientists, activists, and policy-makers).

Trends in the organization of scientific work

The greatest transformation in the organization of scientific work has been the increased orientation to large-scale projects, or “big science” (Weinberg, 1961). As noted in Price’s (1963) landmark analysis, greater size and complexity of research task is reflected in a higher need for collaboration, at least when measured as the number of authors on publications. This trend toward collaboration is increasing and appears to be independent of discipline, as shown in recent analyses of authorship in biology (Zhang, 1997), information science (Lipetz, 1999), social science (Endersby, 1996), and political science (Fisher, Cobane, Vander Ven, & Cullen, 1998). The place of collaboration in science has become so accepted that in 1997 the then chair of the National Science Board, Richard Zare, wrote in a *Science* editorial that future research progress will demand mechanisms to support mega-collaborations, of the type required to solve critical global problems, such as AIDS (Zare, 1997). In Zare’s formulation, research activity would be organized as a form of “distributed intelligence” where experience and knowledge held by scientists at one location could be easily shared and utilized by scientists elsewhere. Specifically, Zare described an era where “...knowledge is available to anyone, located anywhere, at any time; and in which power, information, and control are moving from

centralized systems to individuals” (Zare, 1997, p. 1047). This model suggests a dramatic revision of the historical organization of science away from “invisible colleges,” where the bulk of new knowledge is created by a small core of elite researchers working among themselves (Crane, 1972; Price & Beaver, 1966) – such that 16% of practicing scientists account for about fifty percent of all publications (Price, 1986). By contrast, the notion of distributed intelligence suggests a mobilization of scientific effort – and a corresponding increase in research output and capacity -- so that a larger fraction of the scientific workforce participates in the creation of new knowledge.

The notion of science as distributed intelligence relies heavily on information technology to overcome barriers of time and space. Specifically, studies of scientists and engineers at work suggest that the amount and quality of interaction with colleagues, particularly spontaneous and informal conversations, is an important predictor of productivity (Kraut, Egidio, & Gallegher, 1990; Kraut, Gallegher, & Egidio, 1988; Fox, 1983; Allen, 1977; Pelz & Andrews, 1966; Hagstrom, 1965; Menzel, 1962). Scientists who are remote from communities of elite and active researchers, then, are at a disadvantage in terms of initiating contact with leading investigators that can lead to deeper collaborations. Therefore, the introduction of tools, such as electronic mail, that facilitate easier communication between scientists at non-elite institutions with those at elite institutions could produce increased involvement by non-elite scientists in cutting edge research. This outcome is one possibility predicted by the “peripherality hypothesis” (Sproull & Kiesler, 1991, p. 95). According to this hypothesis, the introduction of electronic communication may produce universal benefits, may produce differential benefits for those who are relatively advantaged, or may produce differential benefits for those who are relatively disadvantaged (as in the previous scenario comparing non-elite and elite scientists). For example, a number of

studies in business settings have shown how employees of global firms use computer-mediated communication to overcome barriers of remote geographic location (Finholt, Sproull, & Kiesler, 2001; Kraut & Attewell, 1997; Constant, Sproull, & Kiesler, 1996). In these cases, electronic mail, bulletin boards, and mailing lists allowed peripheral employees to have the same access to important activities and information flows as centrally located employees, such as those at headquarters sites.

Evidence from a small number of studies appears to support the idea that electronic mail and other computer-mediated communication do enhance scientific productivity (Walsh & Roselle, 1999; Cohen, 1996; Bruce, 1994; Bishop, 1994; Hesse, Sproull, Kiesler, & Walsh, 1993), and that computer networks seem to support larger and more dispersed collaborations (Walsh & Roselle, 1999; Raefeli, Sudweeks, Konstan, & Mabry, 1998; Orlikowski & Yates, 1994). However there is weaker evidence with respect to differential benefits for non-elite scientists (Walsh & Maloney, 2001; Cohen, 1996). For example, in a survey study of 399 scientists in experimental biology, mathematics, physics and sociology, Walsh and Maloney (2001) found slight evidence that electronic mail use was differentially benefiting peripheral scientists relative to core scientists. That is, e-mail use was universally associated with higher productivity, but not in a way that changed the status of peripheral scientists compared to core scientists. Similarly, Cohen (1996) found little support for the equalizing effect of e-mail use among a sample of academic researchers.

One speculation with respect to the absence of a peripherality effect revolves around features of the Internet that give people discretion about selection of communication partners. That is, proponents of the distributed intelligence concept believe that mechanisms like the Internet will break down barriers among disciplines and institutions by allowing people to act on

preferences to associate with others who are different. However, Van Alstyne and Brynjolfsson (1996) argue that if preferences run counter to this expectation, people may choose to associate and communicate mostly with similar others (i.e., because this requires less effort). In this case, then, use of the Internet will produce increased balkanization rather than increased diversification. Under the balkanization scenario, scientists will interact and collaborate more with geographically distributed scientists, but these distant collaborators will be highly similar on critical dimensions (e.g., status, training, methodology and so on).

The collaboratory concept

Studies cited in the preceding section suggest, at least on preliminary examination, that electronic communication – alone – may not be enough to enable a broader range of collaboration in science. This broader collaboration refers both to increased interaction independent of location as well as the reduction of status barriers, such that distinctions between elite and non-elite scientists become less significant. However, electronic communication combined with better access to critical instruments and data may produce differential benefits. That is, while communication is unquestionably important in fostering and sustaining successful scientific collaborations, joint research work also requires access to specialized equipment and unique datasets. This suggests that a true test of the peripherality hypothesis in science requires elaboration of additional network capabilities, particularly applications that enhance sharing of data and data visualizations, and applications that allow remote use of important instruments and facilities.

One mechanism to achieve enhanced access to data and instruments is the “collaboratory.” First proposed by visionary scientists and computer scientists in the late eighties, a collaboratory is “...a center without walls, in which researchers can perform their

research without regard to physical location – interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries” (p. 19, Wulf, 1989). The term collaboratory is a hybrid of collaborate and laboratory. Hence, elaborations of the collaboratory concept stress the simultaneous need to solve problems of control and operation of instrumentation over the Internet, of access and distribution of datasets, and of convenient and flexible interaction with colleagues. Development of computing technology to support collaboratories has not been guided by a grand plan. Rather, systems have emerged through a combination of prodding by visionaries, appropriation of technology designed for other purposes, and the marketing of low-cost high-performance personal computers.

Early proponents of scientific computing anticipated some of the functions of the collaboratory. For example, Vannevar Bush (1945) explored how computers might be used to help scientists keep pace with the explosion of scientific knowledge. He imagined a machine, called the “memex,” that would allow scientists to access and retrieve data and results from a vast array of scientific publications. Pioneers such as Doug Engelbart wrote in the sixties about the use of computing to support intellectual work and built prototype systems for computer-supported meetings (Engelbart, 1963). The initial practical step on the path to collaboratories occurred with the opening of the first wide area computer network in 1969, called the ARPAnet after the Advanced Research Projects Agency of the Department of Defense which sponsored the development of the network. Although originally designed to share scarce computing resources, the most important function of the ARPAnet became its support for electronic mail between computer science and artificial intelligence researchers (Newell & Sproull, 1982). Throughout the seventies and eighties networking technologies developed further, culminating in the creation

of the Internet in 1985, which created the first worldwide community of on-line users (Lynch & Preston, 1990).

The collaboratory idea appeared as scientists recognized the potential represented by expanding national and international computer networks. The first explicit discussion of collaboratories was at an NSF-sponsored workshop in 1989 convened by Joshua Lederberg and Keith Uncapher. This workshop gave the collaboratory concept visibility within the NSF and other relevant national scientific communities. The report of the workshop outlined a number of specific research priorities, including: enabling infrastructure to support collaboratories, construction of collaboratory testbeds in various scientific disciplines, and studies of the process of collaboration and the use of these testbeds by scientists (Lederberg & Uncapher, 1989). One outcome of the 1989 workshop was a series of further workshops in 1993 sponsored by the Computer Science and Telecommunications Board of the National Research Council to explore the feasibility and utility of collaboratories for three disciplines: molecular biology, physical oceanography, and space physics. These fields were chosen for their heterogeneity in size, style of research, technical sophistication, and traditional sources of support. An important result of this activity was the NRC's (1993) report *National Collaboratories: Applying Information Technology for Scientific Research*. The report called for substantial support to develop, refine, and evaluate the collaboratory concept in realistic settings. The impact of these, and other prototype collaboratories is discussed at length later in this chapter.

Four broad changes since the earliest days of the ARPAnet have created conditions conducive for collaboratory development. First, when the ARPAnet appeared its bandwidth was limited and network use was restricted to those institutions with ARPA projects. Today, even the smallest institutions and the most peripheral scientists can have network connections. Second, in

the early days, network connections were scarce. Today, through the proliferation of personal computing and local area networks, network connections are ubiquitous. Third, early user applications had command line interfaces. Today, most software products have intuitive, graphical interfaces that allow users to perform sophisticated actions without learning obscure command sequences. Finally, while early network use was confined to a small community of computer scientists, contemporary users represent a broad spectrum of scientific disciplines as well as a mass audience from business and the general public.

At a more specific level, nearly two decades of technology evolution have led to a rich variety of computer and network tools for the support of collaborative work. Combinations of these existing tools, with elaboration of some new tools, form the core capabilities that constitute a collaboratory. As shown in Figure 1, derived from Atkins (1993), these capabilities can be defined as technology to link people with people, technology to link people with information, and technology to link people with facilities. Examples of people-to-people technologies include familiar applications, such as electronic mail, and tools for data conferencing, such as Microsoft NetMeeting. Technologies to link people with information have recently experienced tremendous growth in sophistication and use, including the World Wide Web and digital libraries (for a more detailed survey of digital libraries see the chapter by Edward Fox and Shalini Urs in this volume as well as the special issue of *Communications of the ACM*, April 1995, edited by Fox, Akscyn, Furuta, and Leggett). Finally, technologies to link people to facilities include data viewers that display the current modes and status of remote instruments as well as services that provide scientifically critical data. An early effort along these lines was the MOS Implementation System, which allowed very large scale integrated chip designers to access remote fabrication facilities (Lewicki, Cohen, Losleben, & Trotter, 1984).

[FINHOLTFIG01 – About here]

Collaboratory experiences

The previous sections have described broad scientific and technological trends that have led to collaboratory development. This section offers a chronological description of notable operational collaboratories, covering the period 1980 to 2000. The focus here is on those collaboratory efforts that have been or are in use by practicing scientists and that have produced accounts of this use. Table 1 summarizes key features of the collaboratory efforts described below.

[FINHOLTTAB01 – About here]

SCIENCENet, 1980s.

SCIENCENet was a proprietary network service initiated in 1980 to meet the unique information and communication needs of the oceanography community. In return for a monthly fee, SCIENCENet subscribers obtained access to colleagues and to data. For example, SCIENCENet supported project-oriented mailing lists used to coordinate activity across scientists in multiple locations. In addition, SCIENCENet provided infrastructure for the storage and transport of large datasets. A critical feature of SCIENCENet was its global coverage, which permitted logins from 45 countries and the Antarctic, as well as via satellite link from research vessels. SCIENCENet was particularly well-suited to the needs of oceanographers, who frequently collaborate to coordinate data gathering across remote locations, or who monitor arrays of automatic sensors, such as buoys moored in the ocean or sensors installed on the polar ice caps.

Hesse, Sproull, Kiesler and Walsh (1993) conducted a systematic analysis of SCIENCENet use and the relationship of usage to scientific outcomes. Data came from a 1988

survey of 338 SCIENCEnet users. Respondents were stratified by category of use, where frequent users had greater than median levels of use and infrequent users had below median levels of use. Scientists were also classified according to location (inland versus coastal), seniority, and disciplinary affiliation (physical oceanography – the target audience for SCIENCEnet – versus other). Outcomes included scientific productivity, measured as publications, professional recognition, and social integration, measured as the extent of each respondent's social network. Results showed that frequent SCIENCEnet users were the more active, productive scientists. As compared with infrequent users, they worked at more prestigious institutions, received more professional recognition, published more, and knew more oceanographers. Controlling for category of use, network usage was still positively related to publications, recognition, and social integration. Further, for inland scientists, use of SCIENCEnet helped to overcome the disadvantage of a non-coastal location, at least in terms of publications. Perhaps the most interesting observation from Hesse et al.'s study was the finding that SCIENCEnet users reported that SCIENCEnet's operators were very sensitive to the special needs of oceanographers – and therefore worked hard to make the system as useful and invisible as possible, in order to allow users to maximize their focus on conducting research.

Worm Community System, 1990-96.

The Worm Community System (WCS) was initiated by Bruce Schatz (1991) and others in 1990 to meet the needs of biologists studying *c. elegans*, a tiny nematode with desirable scientific properties, such as transparent skin and the ability to be frozen and unfrozen for shipping between labs. At the time WCS was built, the community of scientists studying *c. elegans* included 1400 researchers at over 100 labs. The WCS consisted of set of hypertext-linked resources, which was a novel architecture at a time before widespread use of the World

Wide Web. The WCS included: graphics of the *c. elegans* physical structure; a genetic map; formal and informal research notes (including a newsletter called the *Worm Breeder's Gazette*); directory services; a thesaurus; and a database called *acedb*. Star and Ruhleder (1994) described use of the WCS, based on observations of and interviews with over 100 researchers at 25 labs over a three year period, 1991-94. They found that while respondents described WCS as easy to use and relevant to *c. elegans* research, "...most have not signed on; many have chosen instead to use Gopher {an Internet-based document search and retrieval protocol developed by Alberti et al. (1992.1993)} and other simpler net utilities with less technical functionality" (Star & Ruhleder, 1994, p. 254-255). This finding was disappointing for the WCS developers, particularly given the efforts by the WCS team to ensure adequate feedback from the user community during the development of WCS.

Star and Ruhleder identified several factors in the non-use of the WCS. At a high level, a key problem was that the target audience of biologists had to master relatively complex system installations (e.g., choosing the proper X-windows version) within alien computing environments (e.g., Unix workstations). Lower level problems included information barriers ("where do I download the WCS system from?"), unforeseen consequences (e.g., the difficulty of maintaining a unique operating system, such as Unix, among mostly Macintoshes), and political/cultural issues ("what is the appropriate tradeoff between private and public information"). Specifically, in terms of political/cultural factors, many post-docs were reluctant to share ideas or data via WCS, or similar mechanisms, for fear of being anticipated, or scooped, by others. Star and Ruhleder summarized researchers' experience with the WCS in terms of Bateson's (1972) "double bind" concept – where actions with respect to the WCS were often contradictory. For example, users were torn between the power of the WCS versus the inconvenience of learning

new systems and leaving familiar work environments (i.e., most work was performed from a desktop machine, while WCS-related work had to be done from a special workstation).

Upper Atmospheric Research Collaboratory, 1992-99.

The Upper Atmospheric Research Collaboratory (UARC) was initiated in 1993 to serve the needs of a distributed community of space physicists who were users of instruments located at a National Science Foundation-funded observatory located above the Arctic Circle on the west coast of Greenland. Space physics focuses on the interaction between the earth's atmosphere and magnetosphere and the sun. The best known phenomena studied by space physicists are the northern and southern lights, or the photon emissions associated with charged particles in the ionosphere. The goal of UARC was to provide real-time control of remote instruments used to study these and other upper atmospheric events. In addition, UARC was meant to support communication among geographically distributed colleagues about shared real-time data and to provide access to archived data. Finally, UARC was meant to demonstrate the utility of "user-centered" design, where the system evolved through rich interactions between computer scientists, space physicists, and behavioral scientists – with the latter playing a critical role in determining system requirements and evaluating the system in use (Finholt & Olson, 1997; McDaniel, Olson, & Olson, 1994). The original UARC implementation, which was operational from 1992-98, was built in NeXTStep and required either a NeXT workstation or a workstation configured with the NeXTStep operating system. A later version, based on java applets, was operational from 1995-98, and could be run on any operating system with a java compatible web browser.

Olson et al. (1998) provides a summary of experiences with the UARC system. During the NeXTStep era, UARC succeeded in providing real-time output from instruments in

Greenland, notably an incoherent scatter radar, to a small number of users in Europe and in North America. For the most part, during this early period, UARC was used as a tool for collective viewing of live data and for discussion of the data. Typically, scientists would organize activity in small groups around focal observing intervals, called “campaigns,” much as they did pre-UARC – except that UARC eliminated the need for travel to the observatory in Greenland. A key finding related to early UARC use was the observation that the collaboratory expanded the pool of participants in data gathering sessions, compared with traditional sessions, but the additional participants tended to be relatively passive (McDaniel, Olson, & Magee, 1996). That is, in the collaboratory setting, greater ease of access to research activity provided an opportunity for more people to watch – and to increase their level of participation as needed. Finholt, Lewis, & Mott (1995) found that this feature of UARC supported educational use, to the extent that novice space physicists could “lurk” in the collaboratory and observe experienced scientists at work, much as Lave and Wenger (1991) have described novice workers doing in shared physical settings.

Increased use of the World Wide Web within the space physics community, starting around 1994, influenced a re-design of UARC. The second generation system abandoned the NeXTStep environment, which because it required a separate workstation or special modifications to existing workstations, had become a barrier to adoption. The second generation version of UARC also introduced a more scalable data distribution approach. These changes enabled dramatic new kinds of collaboratory use. First, interest in global scale phenomena forced expansion from the original Greenland site to the entire chain of incoherent scatter radars – now including observatories in Norway, Massachusetts, Puerto Rico and Peru. Scientists also demanded output from spacecraft, including imaging satellites in polar orbit as well as satellites

monitoring the solar wind and the surface of the sun. Second, scientists focused on computational models of the upper atmosphere wanted to view their simulated data side-by-side with observational data – and to discuss differences in real-time with experts in interpretation of the observational data. Finally, scientists discovered that tools developed for real-time data gathering and visualization could be easily adapted for viewing archival data. This produced a need for “retrospective” campaigns where the collaboratory was used to view a significant data interval (e.g., a solar sub-storm) from several corroborating instruments (e.g., radar, satellite, computational models and so forth). Later UARC campaigns, as a result of these changes, had a different character involving many more scientists and institutions. For example, in April, 1997 over fifty scientists from twenty labs logged into UARC over a four day period coinciding with a coronal mass ejection.

While the evolution from the NeXTStep-based system to a java-based system did result in the new kinds of use described above, there were still many problems. The goal of the re-design was to produce a UARC system that would be free of the orphaned NeXTStep environment and that would run on the heterogeneous mix of machines and operating systems used in the space physics community. The reality was that the change to java coincided with rapid evolution of the java language, such that the initial java applet version of UARC ran on beta versions of a limited set of web browsers – and then only on some operating systems. Users were forced to frequently download updated browsers, in some cases not even the main browser in use at a site (e.g., Sun’s HotJava browser versus more popular programs). The cost in extra effort for users severely undermined confidence in the UARC system and tarnished otherwise successful efforts to increase the scale and scope of the collaboratory.

Environmental Molecular Sciences Laboratory Collaboratory, 1993 - present.

The Environmental Molecular Sciences Laboratory (EMSL) Collaboratory was initiated in 1993 at the Pacific Northwest National Laboratory (Bair, 1999; Kouzes, Myers, & Wulf, 1996). The collaboratory capability was developed in parallel with creation of the EMSL's physical facility; a collection of instruments and expertise focused on environmental molecular science. Environmental molecular science, in this case, refers to a molecular-level understanding of the physical, chemical and biological processes that underlie remediation of contaminated soils and groundwater, processing of and disposal of stored waste materials, and human health and ecological effects of exposure to pollutants. The EMSL facility consists of data resources (notably a 20 terabyte robotic tape archive), magnetic resonance instruments, and mass spectrometers. Key elements of the collaboratory include applications to support remote operation of the magnetic resonance and mass spectrometer instruments, as well as an electronic notebook for instrument users to record and retrieve data. In addition, collaboratory users have access to a set of generic collaboration tools, including whiteboards, chat rooms, audio and video conferencing, and application sharing (i.e., remote viewing of a shared screen image). As of 1999, the EMSL Collaboratory had gone through three generations of development, with the current system implemented as a java application.

Schur, Keating, Payne, Valdez, Yates and Myers (1998) summarize user experiences in the EMSL Collaboratory. Schur et al. focused on the 200 researchers targeted as potential users of the collaboratory. They conducted interviews with scientists prior to collaboratory use and then observed collaboratory use by a geographically dispersed research team. Researchers contrasted their current practices with a desired ideal, using a standard paradigm of two to five researchers engaged with an eight hour experimental run. Respondents said that under current

practices, disproportionate amounts of time and attention were sunk into preparation for experiments and reporting of experimental results. Collaboration with colleagues occurred mostly around transitions, such as moving from experimental preparation to actually conducting an experiment, and not when needed the most. By contrast, researchers wished to devote the bulk of their time and attention to analysis and interpretation of experimental results, while streamlining preparation and reporting activities. Throughout all experimental activities, they wanted continuous access to colleagues for consultation and discussion, but on an as needed basis.

Observations of scientists in the collaboratory concentrated on a team conducting a protein structure analysis using a Nuclear Magnetic Resonance (NMR) spectrometer and on a team of intelligence analysts working on detection of nuclear material for purposes of detecting non-proliferation treaties. The observational studies produced a number of key findings. First, scientists co-located with research instruments worried that the collaboratory would reduce them to being instrument technicians – doing the bidding of remote scientists at the expense of their own work. In fact, collaboratory use produced sufficient efficiencies such that the local scientists actually had more time to pursue their own projects. Second, much as scientists wished, collaboratory use did allow them to focus more on analysis and less on details of data collection and transmission (e.g., faxes, file transfers, and e-mail). Third, with experience, collaboratory users moved from using collaborative tools in traditional ways, such as one person presenting to others as in a telelecture, to novel techniques more focused on application sharing and data conferencing (even foregoing video-mediated interaction in favor of data conferencing). Finally, audio and shared cursor movements, such as with telepointers, proved much more important and

useful than video in the collaboratory for signaling the beginning and end of tasks, and for avoiding interruptions and talking over others when speaking.

K-12 collaboratories.

A significant potential use of collaboratories is to introduce elementary and secondary students to authentic research practices, such as the CoVis project's efforts (1992-98) to build electronic notebooks to support access and analysis by student teams of the same kinds of data used by practicing scientists (Gomez, Fishman, & Pea, 1998; Fishman, 2000; Edelson, Pea, & Gomez, 1996), or through access to previously exotic research instruments. Good examples of this latter strategy include classroom use of electron microscopes at the Argonne National Laboratory (Zaluzec, 1998) and the nanoManipulator at the University of North Carolina [described in detail with the other NIH collaboratories] (Jones, Superfine, & Taylor, 1999), undergraduate access to nuclear magnetic resonance spectrometers at the Pacific Northwest National Laboratory (Myers, Chonacky, Dunning, & Leber, 1997), and two prototypes that emerged from the World Wide Laboratory (WWL) project at the University of Illinois. The overall goal of the WWL effort has been to provide remote and automated access to imaging instrumentation, such as electron microscopes and nuclear magnetic resonance imaging (MRI) spectrometers, for teams of geographically distributed researchers (Carragher & Potter, 1999). The project has produced a number of Web-based applications for controlling instruments and viewing data displays, aimed at research scientists. An interesting consequence of the Web-based tool development, however, was the realization that with these new tools, access to various high-powered imaging instruments was now available to anyone with an Internet connection and a Web browser.

This realization led to a proposal to use WWL tools in a prototype effort to allow students in kindergarten through high school classrooms to remotely track the development of a chicken embryo using MRI facilities at the University of Illinois' Beckman Institute. This experiment, called Chickscope (1995-98), involved students from ten classrooms, their teachers, and instrument operators at the University of Illinois during the spring of 1996 (Bruce et al. 1997). Eight of the classrooms were in Champaign-Urbana, one was in a nearby rural county, and one was in South Carolina. The project had two goals: a) determine the impact of Internet access to high-powered scientific instrumentation for science instruction in the K-12 environment; and b) test interactive control of the MRI instrument under diverse and adverse conditions (i.e., low bandwidth connections etc.). In addition to remote control of the MRI instrument, Chickscope users also had an archive of images, a chat room, and a special chat area for input from MRI experts. Over the course of the 21 day incubation period, each classroom was granted two twenty minute observation sessions per week. At the end of the experiment, students and teachers reported that access to the MRI was useful – but required some effort to learn how to interpret the MRI images. Everyone was excited to have an additional modality for learning about chicken embryonic development. Finally, teachers reported that the use of Chickscope had a number of beneficial learning outcomes, including increased ability to compare and contrast data, and improved 3-D spatial reasoning (e.g., required to make sense of the planar sections captured by the MRI field of view as it “sliced” through the egg).

Bugscope (1999-present), a successor project to Chickscope, builds on the earlier project's success by reducing costs, such as instrument time and operator effort, while expanding participation (Potter et al. 2000). The focus of Bugscope is the use of an environmental scanning electron microscope (ESEM) to view insects. An ESEM is a special kind of electron microscope

that allows specimens to be viewed in their natural state, i.e., without a conductive coating.

Classrooms submit proposals for studies they wish to perform and approved projects are given microscope time to analyze their samples. In contrast to Chickscope, classrooms only receive one hour of viewing time, but multiple classrooms can participate in a session. As of February, 2000, thirty classrooms involving 1000 students all over the country had used Bugscope. This expansion was achieved while also reducing costs and demands on instrument time. Bugscope users have been enthusiastic in their support for the facility and there are currently over 100 proposals for new Bugscope projects.

DOE 2000 collaboratories.

The EMSL Collaboratory, described above, is one of several collaboratory testbeds funded by the Department of Energy under the DOE 2000 initiative. These other projects include: the Diesel Combustion Collaboratory (1997-2000) (Pancerella, Rahn, & Yang, 1999); the Materials Microcharacterization Collaboratory (MCC, 1997-2000) (Zaluzec, 1998; Zaluzec, 1997); and the Remote Experiment Environment (REE, 1994-97), a collaboratory to support observation and participation in magnetic fusion energy research involving the DIII-D tokamak experiment (McHarg, Caspar, Davis, & Greenwood, 1999; Caspar et al. 1998). A tokamak is a machine for creating a toroidally shaped magnetic confinement field used to contain the plasma, or very high temperature gases, required to achieve a fusion reaction. Fusion is a kind of nuclear reaction in which two light atomic nuclei combine to form another element with the release of energy. The DIII-D tokamak experiment refers to a specific machine, the largest tokamak in the U.S., located at General Atomics in San Diego, California. REE merits extra attention because of the effort to document users' experiences in this collaboratory.

Because operation of the DIII-D tokamak is expensive, scientists must work together in the planning and operation of experimental runs. Within the REE, remote scientists can view data, interact with colleagues at the tokamak site and elsewhere, and observe – via video and audio – activity at the experiment control center. A test use of the REE involved a number of experiments controlled from Lawrence Livermore National Laboratory, in Livermore, California using the DIII-D tokamak (Bly, Keith, & Henline, 1997). Over the course of the experimental runs, scientists at the two locations were able to successfully coordinate their efforts to accomplish a successful experiment. Reactions to the collaboratory-style experiment varied. Remote participants were enthusiastic about their increased access to activity at the DIII-D tokamak, while local participants sometimes resented the intrusion of the “outsiders.” Remote participants also felt excluded from key cues – such as warning lights and alarms – that were available to local participants.

Engineering Collaboratories.

Many of the early collaboratory projects, summarized above, had to produce applications from scratch – such as UARC, the EMSL collaboratory and so forth. Recently, software producers have identified a market for collaboration tools, such as application sharing and presence awareness, and have released a number of products – some of which have become immediately successful, such as Mirabilis’ ICQ (Mirabilis was acquired by America Online, or AOL), AOL’s Instant Messenger, and Microsoft’s NetMeeting. The cost of these tools, free for download in many cases, combined with their ease of use, has led to wide experimentation in business settings. The most pressing need for collaborative technology is often in engineering contexts where engineers need to confer over drawings and other visual data. While use of tools like NetMeeting, which allows data conferencing over images and over output from applications,

has grown dramatically, there are very few systematic studies of NetMeeting use. Two exceptions include recent analyses of NetMeeting use in aerospace engineering (Mark, Grudin, & Poltrock, 1999) and in software engineering (Finholt, Rocco, Bree, Jain, & Herbsleb, 1998).

Both studies focused on use with geographically distributed teams of engineers, where NetMeeting was used along with conventional phone-based audio conferencing. In the aerospace case, principal findings included: over dependence on a small number of technically savvy users; awkward organization of conversational turn taking; high overhead associated with initiating data conferences; restrictive models of use (e.g., only broadcasting briefing slides and not collaborating over joint work); multi-tasking by data conference participants; and problems of awareness (e.g., not knowing who was who, and not knowing who was present at remote locations). For the software engineers, principal findings included: observation of difficulties reconciling different screen resolutions when sharing screens; awkward organization of turn taking – particularly when transferring control over a shared application; the importance of highly motivated NetMeeting “champions” in getting groups over initial learning curves; and the need to run NetMeeting in the background to allow spontaneous data conferencing sessions.

NIH Collaboratories.

The National Institutes of Health have recently launched two significant collaborative initiatives. The National Center for Research Resources (NCRR) made a series of collaborative supplement awards to research resource awardees during the period 1998-2002. These supplements were designed to take existing shared resources, such as instruments and supercomputer simulations, and enhance access to these resources via the addition of network-based collaboration tools. NCRR collaborative awards included support for work on: structural biology via the BioCoRE, or Biological Collaborative Research Environment (Bhandarkar et al.

1999); advanced microscopy via the CMDA, or Collaboratory for Microscopic Digital Anatomy (Hadida-Hassan, Young, Peltier, Wong, Lamont, & Ellisman, 1999; Young et al. 1996); and the “nanoManipulator,” or a mechanism for remotely steering the head of an atomic force microscope (AFM) -- allowing direct manipulation of nanoscale materials (Sonnenwald, Bergquist, Maglaughlin, Kupstats-Soo, & Whitton, 2001; Jones, Superfine, & Taylor, 1999).

In a parallel effort, the National Cancer Institute is funding a virtual Center for AIDS Research spanning four midwestern universities: Northwestern, Minnesota, Wisconsin, and Michigan. The Great Lakes Regional Center for AIDS Research (GLRCFAR, 1998-2002) combines complementary expertise across the sites in a way that none of the sites, alone, could match. Further, the GLRCFAR is notable because it represents the first attempt to build a collaboratory employing only off-the-shelf components. While the GLRCFAR was initiated in 1998, collaboratory use has already become routinized (Teasley, 2001; Teasley & Jain, 2000). For example, CFAR participants login twice a month for a collaboratory-based seminar series, using PlaceWare. PlaceWare is a web-based tool that simulates a virtual lecture hall, based loosely on the LambdaMOO system developed at Xerox Palo Alto Research Center (Curtis, 1997). MOOs (Multi-user dungeon [MUD] Object Oriented) were originally conceived as a virtual space used for text-based adventure games (derived from Dungeons and Dragons, hence the dungeon reference) – but since expanded to cover implementations like PlaceWare, but also other scientific MUDs and MOOS (e.g., Churchill & Bly, 1999; Van Buren, Curtis, Nichols, & Brundage, 1994; Glusman, 1995). In addition to PlaceWare lectures, GLRCFAR scientists regularly confer via NetMeeting, both to write clinical protocols and proposals, and also to view live output from remote instruments (e.g., electron microscope images of patient tissues).

Finally, the GLCFAR is the virtual home for documents and data relevant to joint work across the four member institutions.

Space Physics and Aeronomy Research Collaboratory, 1998-2001.

The Space Physics and Aeronomy Research Collaboratory (SPARC) is a successor project to UARC. As a follow on effort, SPARC has been able to focus on expanding and improving UARC. From an implementation perspective, SPARC is designed as a “thin client” application. This means that users access all features of the collaboratory through a conventional web browser, rather than through specialized software, as in both generations of UARC. For example, in UARC, the initial NeXTStep system was a barrier to use because the technology was exotic and the subsequent java applet system was too unstable to win user confidence. With SPARC, the interface is familiar – anyone who uses a web browser can get started – and demands on local workstations only involve display of conventional web elements, such as standard graphics formats (e.g., JPEG or GIF). A critical operational difference between UARC and SPARC is that SPARC facilities are available continuously, as opposed to only around campaign intervals with UARC. Finally, SPARC represents a re-orientation of the collaboratory to post-hoc data exploration and analysis, termed an electronic workshop, as opposed to real-time data collection, which was the main emphasis in UARC. SPARC still supports real-time data gathering, but scientists found a greater need for retrospective, group investigation of data and visualizations from intervals of known value and interest.

The collaboratory challenge

In 1993, William Wulf wrote about the “collaboratory opportunity.” He noted that the configuration of technologies, needs, and practices were then coming into alignment to make virtual labs possible. In some ways, developments since the early nineties have exceeded Wulf’s

projections. For example, the explosion in Internet use, driven by the World Wide Web, has effected science just as it has effected other spheres of human activity. Yet, by comparison with the breakout success of the Web (Schatz & Hardin, 1994), collaboratory use has been confined to a much smaller number of users. Even within the space of scientific applications on the Web, collaboratories have been dwarfed by digital libraries and knowledge bases, such as the Los Alamos pre-print server (Ginsparg, 1994), GenBank (Ouellette, 1998) and the Protein Data Bank (PDB) (Berman et al. 2000). For example, just considering the PDB, this resource receives an average of between 60,000 to 100,000 hits per day and currently stores 12,592 different structures (Research Collaboratory for Structural Bioinformatics, 2000). It is important to consider whether the relative modest size and growth of collaboratories, compared to systems like the PDB reflects a failure of the original collaboratory vision. Rather than failure, the experience with collaboratories to date indicates the enormous difficulties of supporting complex group work in virtual settings. Overcoming these difficulties represents the great challenge for the next stage of collaboratory development and use. Meeting this challenge involves both extracting lessons learned from previous collaboratory efforts and solving a number of critical problems at the tricky intersection of technology with individual and group behavior.

Meeting the challenge: Lessons learned

A number of tentative conclusions can be drawn from observation of collaboratories in use. It is helpful to start by examining the impact of collaboratory use on the organization and output of work, specifically the work of scientists – since most collaboratories have been targeted at scientific applications. Across the examples described earlier, it is clear that collaboratories have changed the number and type of participants in scientific work. For example, from the UARC and SPARC cases, relaxing the constraint on travel to Greenland and

other remote observatory sites has expanded the number of potential participants in research tasks, such as data collection. In addition, participants are more diverse, both in terms of experience and expertise. The earlier examples also suggest that collaboratories can increase the pace and efficiency of some scientific tasks. For instance, in the case of the Great Lakes Regional Center for AIDS Research, scientists reported that the use of collaboratory tools dramatically reduced the time required to produce a clinical protocol from weeks to hours. Similarly, among space physicists, monitoring conditions via SPARC has allowed them to more effectively use scarce instrument time by activating instruments only under optimal conditions and not according to a pre-determined schedule.

It is less clear that collaboratories have qualitatively changed scientific work, but there is some suggestive evidence. Specifically, use of collaboratory tools forces reflection on resources that become shared that may have previously been unshared, such as data. In one community of brain researchers this realization produced a formal covenant, signed by scientists as a condition of use of the collaboratory, that specified how community data were to be used. This covenant paid particular attention to protecting the interests of younger researchers to prevent senior researchers from anticipating, or “scooping,” their results. Along the same lines, space physicists using UARC and SPARC articulated “rules of the road” describing how public data were to be used, including rights of first publication and mechanisms for sharing credit, such as to instrument owners. In terms of scientific output, collaboratories seem to produce at least two kinds of changes. First, it becomes much easier to combine theoretical visualizations with visualizations of observational data. This helps bridge the gulf that exists in many fields between theoreticians and experimentalists. The capacity to blur the distinction between computational and physical simulations, for example, is a centerpiece of the National Science Foundation’s

George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES). Within NEES, collaboratory tools will allow researchers to combine distributed physical simulation facilities, such as shaking tables and centrifuges, with computational simulations to produce complex real time models of entire structures, or even entire built complexes, such as cities. Second, collaboratories seem to produce a larger field of view. Among space physicists, for example, the ability to view hundreds of instruments worldwide encourages a more global orientation.

1) use does not need to be constant to provide value (general purpose tool VS specialized instrument analogy); 2) systems that are easily integrated into existing work environments are more readily adopted (standalone application VS browser accessible); 3) some domains of activity are more naturally inclined toward collaboration (data collection VS contemplation and idea formation); 4) long distance collaboration creates new expectations for participants, including altered roles (e.g., operators who must be more responsive, students who guide faculty, senior investigators who must accommodate less experienced participants);

Meeting the challenge: Solving critical problems

Perhaps the most significant barrier to both the design and use of collaboratories is that most group practices and routines assume a shared space. For example, studies of distributed cognition show that people inventively exploit features of the social and physical world as resources for accomplishing tasks and thereby reduce their reliance on mental symbolic manipulations (Hutchins, 1995; Pea, 1993). An illustration of this phenomenon is the use by naval navigators of the “three-scale nomogram,” or a pre-drawn chart where the multiplication, addition and division required to compute speed from distance traveled per unit time are represented as complementary logarithmic scales (p. 201, Hutchins, 1990). Changing the circumstances for collaboration, as in collaboratories, may undermine the effectiveness of the

collaborative process by introducing new demands due to loss of a common physical setting.

One critical new demand in the virtual context is that workers must be explicit about information that is normally tacit when co-located.

For example, scientists seated together at a workstation can unambiguously reference features in a data visualization simply by pointing. In a virtual setting, the same scientists must first ensure that they have each produced the same visualization and then ensure that a specific feature referenced by one is the same feature viewed by the other (e.g., through a specific coordinate system or through reference to unmistakable landmarks). In this scenario, the loss of tacit cues in the virtual setting may mean a greater risk of losing common ground (Clark & Brennan, 1991), where common ground is the shared cognitive understanding that allows collaborators to successfully coordinate their effort to accomplish joint work. At a minimum, then, collaborative collaborations may require more effort, in terms of communicating the additional information required to achieve common ground. A challenge for collaborative developers is producing tools and applications that compensate for the absence of shared setting, such as through so-called WISIWYS interfaces (What I See Is What You See). For instance, in both UARC and SPARC, research has focused on mechanisms for data display and data transport to ensure that what one scientist sees can be seen by other scientists (Lee, Prakash, Jaeger, & Wu, 1996; Hall, Mathus, Jahanian, Prakash, & Rasmussen, 1996).

Given the significance of co-location, and more important, the long development of human practices and behaviors contingent on co-location, it is not surprising that attempts to organize activity in virtual settings are difficult. This is the main point Olson and (2001) make in arguing that “distance matters.” Based on laboratory experiments and empirical observations in the field, Olson et al. conclude that for some tasks, co-location is still essential. Specifically,

when tasks are tightly coupled, that is, dependent on frequent interaction and feedback among collaborators, contemporary communication technologies – such as e-mail, video and audio conferencing, and groupware (e.g., Lotus Notes) – do not provide an adequate substitute for co-location (Olson & Teasley, 1996). In part this failure is attributed to inadequate design and poor infrastructure. For example, accurate gaze detection is a key way that humans impute additional meaning in conversations. Yet, most video conferencing applications offer weak support for this kind of fine-grained detail.

Overcoming the difficulties inherent in virtual interaction, then, is partially a matter of elaborating designs and technologies that make virtual settings more like physical settings. Yet, even if successful designs and technologies are identified – there remain critical barriers to successful virtual collaboration. Olson, Finholt, & Teasley (2000) characterize these additional barriers in terms of collaboration readiness and collaboration technology readiness (see also Olson & Olson, 2001; Sonnenwald, 2000; Sonnenwald & Pierce, 1995). Collaboration readiness refers to the extent that potential collaborators are motivated to work with each other. In terms of collaboratory introduction, success seems to require a positive orientation toward collaboration, either as a result of incentives or as a result of normative practice. For example, in the Great Lakes Regional CFAR case, funding from the CFAR was directly tied to willingness to collaborate, as measured by acceptance and use of collaboration tools. This coercive approach worked to bring otherwise reluctant scientists to use PlaceWare, NetMeeting and so forth, resulting in sufficient critical mass to motivate continued use of these tools. In the case of UARC, initial participants in the collaboratory were selected based on pre-existing collaborations. Additionally, space physics as a field has a history of highly collaborative research.

Collaboration technology readiness refers both to the presence of sufficient technology infrastructure, and to the availability of local technology expertise, both explicit and implicit. For example, Olson et al. (2000, p. 12) describe a progression from applications that require minimum training, such as e-mail, to technologies that require greater investment, such as data conferencing tools. Attempts to leapfrog steps in this progression can produce frustration and resistance. Similarly, attempts to implement sophisticated applications, such as desktop video conferencing, will have a higher probability of success when underlying infrastructure is adequate – in this case, access to high bandwidth network connections. In terms of collaboratory development, important lessons can be drawn from the difficulties described earlier with respect to the evolution of UARC from a NeXTStep-based system to a java-based system. That is, apparent technological advantages, such as the purported universality of java code, needed to be weighed against equally important factors of familiarity and reliability.

Future collaboratory development

Laboratories emerged as physical settings designed to house rare and expensive instruments, and the scientists using the instruments. The forms of social organization that grew out of this arrangement depended heavily on co-location. Today, the evolution of information technology suggests a form of collaboration without proximity. Specifically, the goal of collaboratory development is the creation of “laboratories without walls.” This concluding section explores the consequences for scientific practice and for scientific communities when collaboration becomes independent of physical location.

As noted earlier, in discussion of the peripherality hypothesis, a hope for collaboratory elaboration and use is that improved access to important, but scarce instruments and data, combined with easy communication among researchers will diminish barriers of status, time, and

space that hamper scientific progress. However, it is important to note that powerful forces will continue to exist that will move collaboratories in the direction of exclusivity and selection that have characterized the historic organization of science. First, the availability of a means for contact between two scientists does not guarantee that contact will occur. For instance, science in the virtual realm may be just as likely as traditional science to be typified by strict enforcement of boundaries defining invisible colleges. In an examination of an early system that supported network mediated communication among scientists, Hiltz and Turoff (1993) found that elite scientists using the system were more likely to receive messages than non-elite scientists, but that elite scientists were more likely to ignore the messages they received, particularly when those messages were sent by non-elite scientists. Second, economic considerations dictate that some scientific data and results will always be secured from widespread access. In chemistry, the bulk of practicing chemists are employed in private firms. These firms have proprietary interests in the products of their employees, specifically intellectual property such as patentable compounds and processes. As a result, chemists as a group use public computer networks less than other scientific disciplines that are dominated by academic practitioners (Walsh & Bayma, 1996). Third, scientific collaborations appear to require face-to-face contact, at least initially, suggesting that conferences and invited meetings will continue to function as critical filters on scientific participation. In a study of interpersonal communication networks among computer scientists Carley and Wendt (1991) found that face-to-face contact was critical in starting a scientific relationship. While the computer scientists used email to maintain existing collaborations, none of the identified collaborative relationships started via email.

Collaboratory advocates envision the ultimate withering away of physical laboratories. However, it seems more realistic to suggest that collaboratory use will augment, but not replace,

proximity as a tool for fostering scientific collaboration. Further, the benefits of collaboratory use may differ by the status and experience of collaboratory users. Opportunities and gains seem most obvious for graduate and undergraduate students and non-elite scientists, since these are often the members of the scientific community least able to travel and meet other scientists. Collaboratories may represent a mechanism for accelerating students' immersion into important networks. For example, through UARC, space physics graduate students were able to participate in experiments during their first year, while in the past this did not occur until the third or fourth year. For elite scientists, collaboratories may offer more imposition than benefit. Specifically, if collaboratory sessions become opportunities for non-elite scientists and students to bombard these senior investigators with questions or demands, the senior scientists may respond by withdrawing their participation (and rely on traditional means for continuing collaborations). Finally, for non-elite scientists collaboratories may provide broader access to some resources, such as instrument time, and may deliver access to elite scientists (although still at the discretion of the elite scientists). Most important, non-elites may use collaboratories to foster links among one another, which could be both valuable and damaging (vis the balkanization outcome described by Van Alstyne and Brynjolfsson (1996) – where the Internet may reinforce connections among those who are similar versus creation of more diverse links). From the perspective of creating an intellectual community, the collaboratory may fill a critical niche, particularly for scientists at smaller institutions where they may have few local colleagues. However, if the concentration of non-elites is taken as an indication of the secondary status of a community, collaboratories may become the home for scientists who are marginalized in their larger, more traditional scientific communities. An instance of this may be the phenomenon of

“e-journals,” which are numerous on the Web (Odlyzko, 1999) – yet continue to have a clearly inferior status relative to traditional journals.

In summary, the emergence of collaboratories represents an important convergence of computing technology with scientific practice. Collaboratories, by themselves, will not produce changes in science. However, at this early stage in their development, it may be possible to anticipate openings for change afforded by collaboratories and be prepared to exploit these openings. This means that those in the scientific community, and beyond, should actively explore how collaboratories can be used to expand participation in science, rather than accepting collaboratories and other new technologies as extensions of the status quo.

Bibliography

Alberti, B., Anklesaria, F., Lindner, P., McCahill, M., and Torrey, D. (1992). The internet gopher protocol: a distributed document and search retrieval protocol [Machine-readable data file]. Microcomputer and Workstation Networks Center (Producer). University of Minnesota Gopher Server (Distributor)

Allen, T.J. (1977). *Managing the flow of technology*. Cambridge, MA: MIT Press.

Anklesaria, F., McCahill, M., Linder, P., Johnson, D., Torrey, D., & Albert, B. (1993). The Internet Gopher Protocol (a distributed document search and retrieval protocol). RFC 1436.

Atkins, D.E. (1993, December). *Overview of the Upper Atmospheric Research Collaboratory*. Paper presented at the Upper Atmospheric Research Collaboratory Workshop, Menlo Park, CA.

Bair, R. A. (1999). Collaboratories: Building electronic scientific communities. In T.H. Dunning (Chair), *Impact of advances in computing and communications technologies on chemical science and technology: Report of a workshop* (pp. 125-140). Washington, D.C.: National Academy Press.

Bateson, G. (1972). *Steps to an ecology of mind*. New York: Ballantine.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., & Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28, 235-242.

Bhandarkar, M., Budescu, G., Humphrey, W.F., Izaguirre, J.A., Izrailev, S., Kale, L.V., Losztin, D., Molnar, F., Phillips, J.C., & Schulten, K. (1999). BioCoRE: A collaboratory for structural biology. In A. G. Bruzzone, A. Uchrmacher, & E. H. Page (Eds.), *Proceedings of the*

SCS International Conference on Web-Based Modeling and Simulation (pp. 242-251). San Francisco, CA.

Bishop, A.P. (1994). The role of computer networks in aerospace engineering. *Library Trends*, 42, 694-729.

Bly, S., Keith, K.M., & Henline, P.A. (1997). *The work of scientists and the building of collaboratories*. General Atomics Report #GA-A22619.

Bruce, B.C., Carragher, B.O., Damon, B.M., Dawson, M.J., Eurell, J.A., Gregory, C.D., Lauterbur, P.C., Marjanovic, M.M., Mason-Fossum, B., Morris, H.D., Potter, C.S., & Thakkar, U. (1997). Chickscope: An interactive MRI classroom curriculum innovation for K-12. *Computers and Education Journal*, 29(2), 73-87.

Bruce, H. (1994). Internet services and academic work: An Australian perspective. *Internet Research*, 4, 24-34.

Bush, V. (1945). As we may think. *The Atlantic Monthly*, 176, 101-108.

Carley, K., & Wendt, K. (1991). Electronic mail and scientific communication: A study of the SOAR extended research group. *Knowledge: Creation, Diffusion, Utilization*, 12, 406-440.

Carragher, B., & Potter, C.S. (1999). The World Wide Laboratory: Remote and automated access to imaging instrumentation. In T.H. Dunning (Chair), *Impact of advances in computing and communications technologies on chemical science and technology: Report of a workshop* (pp. 141-153). Washington, D.C.: National Academy Press.

Caspar, T.A., Meyer, W.M., Moller, J.M., Henline, P., Keith, K., McHarg, B., Davis, S., & Greenwood, D. (1998). Collaboratory operations in magnetic fusion research. *ACM Interactions*, 5(3), 56-64.

Churchill, E.F., & Bly, S. (1999). It's all in the words: Supporting work activities with lightweight tools. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work, Group '99* (pp. 40-49). New York: ACM Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially-shared cognition* (pp. 127-149). Washington, DC: American Psychological Association.

Cohen, J. (1996). Computer mediated communication and publication productivity among faculty. *Internet Research*, 6(2-3), 41-63.

Constant, D., Sproull, L.S., & Kiesler, S.B. (1996). The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organizational Science*, 7, 199-135.

Crane, D. (1972). *Invisible colleges*. Chicago: University of Chicago Press.

Curtis, P. (1997). Mudding: Social phenomena in text-based virtual realities. In S. Kiesler (Ed), *Culture of the internet*,(pp. 121-142). Mahwah, NJ: Lawrence Erlbaum Associates.

Edelson, D.C., Pea, R.D., and Gomez, L.M. (1996). The collaboratory notebook. *Communications of the ACM*, 39(4), 32-33.

Endersby, J.W. (1996). Collaborative research in the social sciences: Multiple authorship and publication credit. *Social Science Quarterly*, 77(2), 375-392.

Engelbart, D.C. (1963). A conceptual framework for the augmentation of man's intellect. In P.D. Howerton & D.C. Weeks (Eds.) *Vistas in information handling, Volume 1* (pp. 1-29). Washington, D.C.: Spartan Books.

Finholt, T.A., and Olson, G.M. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, 8(1), 28-36.

Finholt, T.A., Lewis, S.A., & Mott, W.H. (1995). Distance learning in the Upper Atmospheric Research Collaboratory. Unpublished manuscript. Collaboratory for Research on Electronic Work, The University of Michigan, Ann Arbor, MI.

Finholt, T.A., Rocco, E., Bree, D., Jain, N., & Herbsleb, J. (1998). NotMeeting: A field trial of NetMeeting in a geographically distributed organization. *SIGGROUP Bulletin*, 21(1), 66-69.

Finholt, T.A., Sproull, L., & Kiesler, S. (in press). Outsiders on the inside: Sharing know-how across space and time. In P. Hinds & S. Kiesler (Eds.), *Distributed work*. Cambridge, MA: MIT Press.

Fisher, B.S., Cobane, C.T., Vander Ven, T.M., & Cullen, F.T. (1998). How many authors does it take to publish an article? Trends and patterns in political science. *PS-Political Science and Politics*, 31(4), 847-856.

Fishman, B. (2000). How activity fosters CMS tool use in classrooms: Reinventing innovations in local contexts. *Journal of Interactive Learning*, 11(1), 3-27.

Fox, E.A., Akscyn, R.M., Furuta, R.K., & Leggett, J.J. (Eds.). (1995). Digital libraries [Special issue]. *Communications of the ACM*, 38(4).

Fox, M.F. (1983). Publication productivity among scientists. *Social Studies of Science*, 13, 285-305.

Ginsparg, P. (1994). First steps toward electronic research communication. *Computers in Physics*, 8(4), 390-396.

Glusman, G. (1995). *BioMOO's purpose* [On-line]. Available: <http://bioinformatics.weizmann.ac.il:70/0/biomoo/purpose>

- Gomez, L., Fishman, B., & Pea, R. (1998). The CoVis project: Building a large-scale science education testbed. *Interactive Learning Environments*, 6(1-2), 59-92.
- Hadida-Hassan, M., Young, S.J., Peltier, S.T., Wong, M., Lamont, S., & Ellisman, M.H. (1999). Web-based telemicroscopy. *Journal of Structural Biology*, 125, 235-245.
- Hagstrom, W.O. (1965). *Scientific community*. New York: Basic Books.
- Hall, R.W., Mathur, A.G., Jahanian, F., Prakash, A., & Rasmussen, C. (1996). Corona: A communication service for scalable, reliable group collaboration systems. In *Proceedings of the ACM 1996 Conference on Computer-supported Cooperative Work* (pp. 140-149). New York: ACM Press.
- Hesse, B. W., Sproull, L.S., Keisler, S.B., & Walsh, J.P. (1993). Returns to science: Computer networks in oceanography. *Communications of the ACM*, 36(8), 90-101.
- Hesse, B.W., Sproull, L.S., Kiesler, S., & Walsh, J.P. (1993). Returns to science. *Communications of the ACM*, 36, 90 101.
- Hiltz, S.R., & Turoff, M. (1993). *The network nation: Human communication via computer* (Rev. ed.). Cambridge, MA: MIT Press.
- Hutchins, E. (1990). The technology of team navigation. In J. Galegher, R.E. Kraut, & C. Egido (Eds.) *Intellectual teamwork: Social and technological foundations of cooperative work* (pp. 191-220). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hutchins, E. (1995) *Cognition in the wild*. Cambridge, MA: MIT press.
- Jones, M.G., Superfine, R., & Taylor, R. (1999). Virtual viruses. *Science Teacher*, 66(7), 48-50.
- Katz, J.S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31-43.

Kouzes, R.T., Myers, J.D., & Wulf, W.A. (1996). Collaboratories: Doing science on the Internet. *IEEE Computer*, 29(8), 40-46.

Kraut, R.E., & Attewell, P. (1997). Media use in a global corporation: Electronic mail and organizational knowledge. In S. B. Kiesler (Ed.), *Culture of the Internet* (pp. 323-342). Mahwah, NJ: Lawrence Erlbaum Associates.

Kraut, R.E., Egidio, C., & Galegher, J. (1990). Patterns of contact and communication in scientific research collaboration. In R. Kraut, C. Egidio, and J. Galegher (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work*, (pp. 149-171). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kraut, R.E., Galegher, J., & Egidio, C. (1988). Relationships and tasks in scientific collaboration. *Human-Computer Interaction*, 3, 31-58.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Lederberg, J., & Uncapher, K. (1989). *Towards a national collaboratory: Report of an invitational workshop at the Rockefeller University, March 17-18*. Washington, D.C.: National Science Foundation, Directorate for Computer and Information Science Engineering.

Lee, J.H., Prakash, A., Jaeger, T., & Wu, G. (1996). Supporting multi-user, multi-applet workspaces in CBE. *Proceedings of the ACM 1996 Conference on Computer-Supported Cooperative Work* (pp. 344-353). New York: ACM Press.

Lewicki, G., Cohen, D., Losleben, P., & Trotter, D. (1984). MOSIS: Present and future. In P. Penfield (Ed.), *Advanced research in VSLI* (pp. 124-128). Dedham, MA: Artech House, Inc.

- Lipetz, B.A. (1999). Aspects of JASIS authorship through five decades. *Journal of the American Society for Information Science*, 50(11), 994-1003.
- Lynch, C., & Preston, C. (1990). Internet access to information resources. *Annual Review of Information Science and Technology*, 25, 263-312.
- Mark, G., Grudin, J., & Poltrock, S.E. (1999). Meeting at the desktop: An empirical study of virtually collocated teams. In *Proceedings of the Sixth European Conference on Computer-Supported Work '99* (pp. 159-178). Dordrecht, NL: Kluwer Academic.
- McDaniel, S.E., Olson, G.M., & Magee, J. (1996). Identifying and analyzing multiple threads in computer-mediated and face-to-face communication. In *Proceedings of CSCW '96* (pp. 39-47). New York: ACM Press.
- McDaniel, S.E., Olson, G.M., & Olson, J.S. (1994). Methods in search of methodology: Combining HCI and object orientation. In *Proceedings of CHI '94* (pp. 145-151). New York: ACM Press.
- McHarg, B.B., Casper, T.A., Davis, S., and Greenwood, D. (1999). Tools for remote collaboration on the DIII-D national fusion facility. *Fusion Engineering and Design*, 43(3-4), 343-355.
- Menzel, H. (1962). Planned and unplanned scientific communication. In B. Barber & W. Hirsch (Eds.), *The sociology of science* (pp. 417-441). Glencoe, IL: Free Press.
- Myers, J.D., Chonacky, N., Dunning, T., & Leber, E. (1997). Collaboratories: Bringing national laboratories into the undergraduate classroom and laboratory. *Council on Undergraduate Research Quarterly*, 17(3), 116-120.
- National Research Council (1993). *National collaboratories: Applying information technology for scientific research*. Washington, D.C.: National Academy Press.

Newell, A., & Sproull, R.F. (1982). Computer networks: Prospects for scientists. *Science*, 215, 843-852.

Odlyzko, A. (1999). The economics of electronic journals. In R. Ekman and R. Quandt (Eds.), *Technology and scholarly communication* (pp. 380-393). Berkeley: University of California Press.

Olson, G.M., Atkins, D.E., Clauer, R., Finholt, T.A., Jahanian, F., Killeen, T.L., Prakash, A., & Weymouth, T. (1998). The Upper Atmospheric Research Collaboratory (UARC). *ACM Interactions*, 5(4), 48-55.

Olson, G.M., Finholt, T.A., & Teasley, S.D. (2000). Behavioral aspects of collaboratories. In S.H. Koslow & M.F. Huerta (Eds.), *Electronic collaboration in science* (pp. 1-14). Mahwah, NJ: Lawrence Erlbaum Associates.

Olson, G.M., & Olson, J.S. (2001). Distance matters. *Human Computer Interaction*, 15, 139-179.

Olson, J.S., & Teasley, S.D. (1996). Groupware in the wild: Lessons learned from a year of virtual collocation. In *Proceedings of the ACM 1996 Conference on Computer-Supported Work* (pp. 419-427). New York: ACM Press.

Orlikowski, W.J., & Yates, J. (1994). Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly*, 39, 541-574.

Ouellette, B.F. (1998). The GenBank sequence database. *Methods of Biochemistry Analysis*, 39, 16-45.

Pancerella, C.M., Rahn, L.A., and Yang, C.L. (1999, November). *The Diesel Combustion Collaboratory: Combustion researchers collaborating over the Internet*. Paper presented at Supercomputing '99, Portland, OR.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 47-87). Cambridge: Cambridge University Press.

Pelz, D.C., & Andrews, F.M. (1966). *Scientists in organizations*. New York: John Wiley & Sons.

Potter, C.S., Carragher, B.O., Carroll, L., Conway, C., Grosser, B., Hanlon, J., Kisseberth, N., Robinson, S., Stone, D., Thakkar, U., & Weber, D. (2000). Bugscope: A sustainable microscope project for K-12. In B. Fishman & S. O'Connor-Divelbiss (Eds.) *Fourth International Conference on the Learning Sciences* (pp. 376-383). Mahwah, NJ: Lawrence Erlbaum Associates.

Price, D.J. de S. (1963). *Little science, big science*. New York: Columbia University Press.

Price, D.J. de S. (1986). *Little science, big science...and beyond*. New York: Columbia University Press.

Price, D.J. de S., & Beaver, D.D. (1966). Collaboration in an invisible college. *American Psychologist*, 21(11), 1011-1018.

Raefeli, M.S., Sudweeks, F., Konstan, J., & Mabry, E. (1998). ProjectH overview: A collaborative quantitative study of computer-mediated communication. In F. Sudweeks, M. McLaughlin, M.S. Rafaeli (Eds.), *Network and netplay: Virtual groups on the Internet* (pp. 265-282). Cambridge, MA: MIT Press.

Research Collaboratory for Structural Bioinformatics (2000). *Protein Data Bank Annual Report, 1999-2000*. New Brunswick, NJ: Rutgers University.

Schatz, B. (1991-92). Building an electronic community system. *Journal of Management Information Systems*, 8(3), 87-101.

Schatz, B.R., & Hardin, J.B. (1994) NCSA Mosaic and the World Wide Web: Global hypermedia protocols for the Internet. *Science*, 265, 895-901.

Schur, A., Keating, K.A., Payne, D.A., Valdez, T., Yates, K.R., & Myers, J.D. (1998). Collaborative suites for experiment-oriented scientific research. *ACM Interactions*, 3(5), 40-47.

Sonnenwald, D.H. (1995). Contested collaboration: A descriptive model of intergroup communication in information system design. *Information Processing and Management*, 31(6), 859-877.

Sonnenwald, D.H., Bergquist, R., Maglaughlin, K.A., Kupstas-Soo, E., & Whitton, M. (2001). Designing to support collaborative scientific research across distances: The nanoManipulator example. In E. Churchill, D. Snowdon, A. Munro (Eds.), *Collaborative Virtual Environments* (pp. 202-224). London: Springer Verlag.

Sonnenwald, D.H., & Pierce, L. (2000). Information behavior in dynamic group work contexts: Interwoven situational awareness, dense social networks and contested collaboration in command and control. *Information Processing and Management*, 36(3), 461-479.

Sproull, L.S., & Kiesler, S.B. (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: MIT Press.

Star, S.L., & Ruhleder, K. (1994). Steps toward an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. In *Proceedings of the Conference on Computer-Supported Cooperative Work 1994* (pp. 253-264). New York: ACM Press.

Teasley, S.D., & Jain, N. (2000). Scientists together working apart: The CFAR collaboratory. Unpublished manuscript. Collaboratory for Research on Electronic Work, The University of Michigan, Ann Arbor, MI.

Teasley, S.D. (2001, February). Yours, mine and ours: Sharing data in geographically distributed collaborations. In B. Jasny & P.J. Hines (Chairs), *Staying afloat in a sea of data*. Symposium conducted at the 167th National Meeting of the American Association for the Advancement of Science, San Francisco, CA.

Traweek, S. (1992). *Beamtimes and lifetimes: The world of high energy physics*. Cambridge, MA: Harvard University Press.

Van Alstyne, M., & Brynjolfsson, E. (1996). Could the Internet balkanize science? *Science*, 274, 1479-1480.

Van Buren, D., Curtis, P., Nichols, D. A., and Brundage, M. (1994, September). *The AstroVR collaboratory: An on-line multi-user environment for research in astrophysics*. Paper presented at the Fourth Annual Conference on Astronomical Data Analysis Software and Systems, Baltimore, MD.

Walsh, J.P., & Bayma, T. (1996). Computer networks and scientific work. *Social Studies of Science*, 26, 661-703.

Walsh, J.P., & Maloney, N. (in press). Scientific productivity, new communication technology and inequality: Does CMC narrow the gap? In P. Hinds & S. Kiesler (Eds.), *Distributed work*. Cambridge, MA: MIT Press.

Walsh, J.P., & Roselle, A. (1999). Computer networks and the virtual college. *Encyclopedia of Library and Information Science*, 65, 22-42.

Weinberg, A. (1961). Impact of large-scale science on the United States. *Science*, *134*, 161-164.

Wulf, W.A. (1989). The national collaboratory – a white paper. In J. Lederberg & K. Uncaphar, *Towards a National Collaboratory: Report of an invitational workshop at the Rockefeller University, March 17-18* (Appendix A). Washington, D.C.: National Science Foundation, Directorate for Computer and Information Science Engineering.

Wulf, W.A. (1993). The collaboratory opportunity. *Science*, *261*(5123), 854-855.

Young, S.J., Fan, G.G.Y., Hessler, D., Lamont, S., Elvins, T.T., HadidaHassan, M., Hanyzewski, G.A., Durking, J.W., Hubbard, P., Kindlmann, G., Wong, E., Greenberg, D., Karin, S., and Ellisman, M.H. (1996). Implementing a collaboratory for microscopic digital anatomy. *International Journal of Supercomputer Applications and High Performance Computing*, *10*(2-3), 170-181.

Zaluzec, N. (1997, October). *Interactive collaboratories for microscopy and microanalysis*. Paper presented at the Multinational Conference on Electron Microscopy-97, Portoroz, Slovenia.

Zaluzec, N. (1998). The telepresence microscopy collaboratory. *Teleconference*, *17*(9), 17-20.

Zare, R.N. (1997). Knowledge and distributed intelligence. *Science*, *275*, 1047.

Zhang, H.Q. (1997). More authors, more institutions, and more funding sources: Hot papers in biology from 1991 to 1993. *Journal of the American Society for Information Science*, *48*(7), 662-666.

Table 1

Descriptive characteristics of U.S. collaborative efforts, 1992 – 2000

Project title	Start date	Sponsor	Total budget	Target community	Peak use	Total use	Relevant studies
Worm Community System	1990	NSF	\$1,741,141	<i>c. elegans</i> researchers	NA	NA	Star94
UARC	1992	NSF	\$4,455,329	Space physicists	54	163	Finholt97
CoVis	1992	NSF	\$4,794,633	K-12 students	9,225	20,250	Gomez98
CMDA	1994	NSF/NIH	\$2,273,321	Electron microscopy	NA	NA	Hadida-Hassan99
Remote Experimental Environment	1995	DOE	NA	Fusion researchers	NA	NA	Caspar98
EMSL Collaboratory	1995	DOE	NA	NMR users	5	17	Bair99
Chickscope	1996	U. of Ill.	NA	Microscopy, K-12 students	30	900	Bruce97
MMC Collaboratory	1997	DOE/NIST	\$10,890,000	Electron microscopy	6	80	Zaluzec98
Diesel Combustion Collaboratory	1997	DOE	\$7,155,000	Diesel combustion	NA	NA	Pancerella99
Great Lakes Regional CFAR	1998	NIH	\$814,088	AIDS researchers	25	86	Teasley01
SPARC	1998	NSF	\$2,440,000	Space physicists	34	215	Olson00
BioCoRE	1998	NIH	\$1,225,000	Structural biologists	NA	NA	Bhandarkar99
NanoManipulator	1999	NIH	\$1,280,000	Remote use of AFM	2	114	Sonnenwald01
Bugscope	1999	NSF	\$447,751	Microscopy, K-16 students	30	1500	Potter00

Figure captions

Figure 1. The collaboratory concept: Using distributed, media-rich network connections to link people to each other, to facilities, and to information

