ITR/SOC + IM: Sustainable and Generalizable Technologies to Support Collaboration in Science

Since science has always been a form of what we now call "distributed knowledge work", scientists were among the first to recognize the potential of emerging information and communication technologies. For instance, electronic mail first became widespread within scientific subcommuniities. As additional networked tools became available a more coherent vision has emerged of how technology-mediated science can be conducted. By the late 1980s the concept of a collaboratory was being discussed at places like the National Science Foundation and the National Research Council. Collaboratories were defined as a "…'center without walls,' in which the nationals researchers can perform their research without regard to geographical location [Wulf, 1989]." The vision was that scientists who are geographically dispersed could work together using appropriate technology to access each other, remote tools, databases, and instruments (National Research Council, 1993; Kouzes, Myers & Wulf, 1996; Finholt & Olson, 1997).

Over the past decade there have been a series of collaboratory projects funded by NSF, DOE, NIH, and other agencies, some successful and some less so. These projects provide us with a base of experience from which we have begun to form generalizations about the conditions for success. These projects have demonstrated the promise of the vision. Indeed, it is feasible and useful to use networks to link teams of people, data, tools, and facilities to reduce the barriers of time and distance.

However, the design, deployment, and adoption of new collaboratories remain difficult and uncertain processes. Each collaboratory has been built as an independent effort. Since these efforts involved complex responses to often idiosyncratic mixtures of social and technical factors, general lessons about collaboratory design remain elusive. The large effort required to produce the first prototype collaboratories has not allowed careful reflection about broader principles of collaboratory development. These principles are needed to expand collaboratory use beyond narrow application in a few scientific fields.

We seek to change this. We aim to define, abstract, and codify the broad underlying technical and social elements that lead to successful collaboratories. We seek to synthesize the vocabulary, associated principles, design methods, and technical infrastructure for propagating and sustaining collaboratories across a wide range of circumstances. Our goal is for users with a need for collaboratory infrastructure to be able to create successful collaboratories on their own. We believe this goal can be achieved within the five year scope of our proposed project. An even more ambitious goal would be to have collaboration capabilities become integrated into the common infrastructure that any scientist could access simply by being a practicing member of a relevant community. This more ambitious goal is outside the scope of the present project, but it is a vision that drives our work.

We will pursue this overall goal through a series of activities associated with six specific objectives. Each of these is described in greater detail later in this proposal.

Specific Objectives

(1) The qualitative and quantitative study of **collaboratory design and usage**, examining both technical and social aspects of performance. We will convene a series of **workshops** with a broad spectrum of collaboratory developers/evaluators, and systematically explore their experiences to develop generalizations.

(2) Creation and maintenance of a **Collaboratory Knowledge Base**. Drawing on the work in (1), we will create a Web-accessible archive of primary source material, summaries and abstracts, and relevant generalizations and principles, a database of collaboratory resources, and other related material.

(3) The abstraction and codification of **principles, heuristics, and frameworks** to guide the rapid creation and deployment of successful collaboratories. We seek to understand the complex interplay of technical and social factors that can lead to **principles of design or customization.** Codification of these principles will reduce the dependence of scientific collaborators on specialized collaboratory developers for building and deploying such tools. We also seek deeper understanding of the

sociotechnical processes involved with collaboratory construction and use, both as an element of improving collaboratories and as a basic research goal.

(4) The formulation of a **components approach** to collaboratory technology and guidelines for their use. Many collaborative technologies already exists, but not in a form that allows for easy configuration and customization by end users. We will develop technical specifications for such components, and work with commercial partners or open source developers to see that such components emerge. Our goal here is to lower the technical as well as the social barriers to participation in collaboratories.

(5) Development of **new collaboratory testbeds**. To test our principles and methods, we will develop and deploy several new testbed projects in areas that require new technical capabilities and that increase our understanding of the social processes behind successful collaboration. We will particularly explore the **knowledge transfer process to end users** by overseeing at least one community in the process of assembling their own collaboratory infrastructure.

(6) The development of a **collaboratory curriculum** that will be used to train domain specialists in how to adapt collaboratory tools and principles to their science. Drawing on our lessons and generalizations, we will formulate a curriculum and offer it through a special **certificate program**. The goal is to accelerate the expansion of collaboratory principles into all of the sciences through the training of domain scientists.

The Michigan Experience Base

The proposed research will involve empiricism, abstraction, and application, requiring an interdisciplinary approach. The School of Information at the University of Michigan is uniquely poised for such a research challenge, with a decade of experience in the development and evaluation of collaboratories. The faculty is drawn from a wide range of disciplines including computer science, information science, human-computer interaction, psychology, economics, organizational theory, system design, and electronic archiving. It has an experienced professional staff to support the design, assembly, deployment, and evaluation of experimental systems. The proposed research focuses this interdisciplinary strength toward a unifying goal of creating both principles and design methods for collaboratory development. We seek to bring together researchers, students, and partners from business and non-profits to help nurture an emerging R&D community that will generate and propagate knowledge about the design and use of collaboratories.

Experience has shown that collaboratory development is a microcosm of social change in general. The factors that can play a major role cross the full spectrum of social and technical fields. Collaboratory projects can fail because of insufficient data structures, misaligned incentives to participate, non-scalable network technologies, users' cognitive limitations, a lack of technical standard sufficient to facilitate interoperability, and cultural incompatibilities among participants. These examples suffice to illustrate that no single theoretical perspective or disciplinary competence will cover all the issues that must be addressed in successful collaboratory building.

While a wide range of disciplines are necessary, at the same time, codifying methods for building collaboratories requires easy and efficient movement of ideas among specialists in the multitude of relevant fields. It requires diverse competencies, as well as common knowledge and trust.

The researchers in this proposal are unusual both in the diversity of their fields of expertise, and in the depth of their experience in surfacing and building upon each other's insights. Nearly all the researchers have significant affiliation with the University of Michigan's School of Information. Many have worked together for 5 years-- in some cases more than 10 years -- on projects that required just this kind of "coherent diversity." They have not only written papers together, but have also built and deployed systems, taught together and designed successful joint curricula at the masters and doctoral levels. This diversity of backgrounds around a core of shared concepts and team-based experience is a crucial asset in a project as broad in scope as the one we propose.

The following is a brief list of the collaboratory projects conducted at the University of Michigan by the participants in this proposal:

Science collaboratories:

SPARC/UARC. The Space Physics and Aeronomy Research Collaboratory (formerly known as the Upper Atmospheric Research Collaboratory) has been funded continuously by NSF since 1992. The project has focused on supporting real-time access to remote data sources, for the purpose of carrying out collaborative scientific campaigns (Finholt & Olson, 1997; Olson et al., 1998). We provide real-time access to hundreds of data sources worldwide, both land-based and satellite. In addition, we provide access to supercomputer models so comparisons between data and theory can be made in real time. Dozens of scientists conduct real-time scientific conversations over a chat facility as phenomena in the upper atmosphere unfold. Overall, hundreds of scientists from around the world have participated. We are currently extending this to on-line workshops to support scientific activity between observational campaigns, with access to databases, analysis tools, and digital libraries.

Great Lakes Center for Aids Research (CFAR). This project, funded for five years by NIH in 1998, supports AIDS/HIV researchers at Michigan, Northwestern, Wisconsin, and Minnesota. Researchers with different expertise are able to view and discuss data (e.g., images of tissue sample) and negotiate new study protocols. They also use distance tools to hold cross-site lab meetings, and share colloquia speakers. These activities have accelerated the progress of their research.

Brain Science Collaboratory. This project started in 1998 funded by the Pritzker Foundation, and links mood disorder researchers at Michigan, Stanford, and Cornell to coordinate clinical trials, functional MRI imaging, and bench science. Selected Internet-based collaboration tools have been introduced, though technical complexities such as network heterogeneity have slowed the adoption of new practices.

Medical Collaboratory. NSF funded this project between 1994 and 1997. It focused on remote access to radiological images, and developed asynchronous tools to allow general practitioners and radiologists to interact over such images. Extensive studies of clinical practice were done prior to deploying a prototype system that linked a family practice site with a university hospital (Yakel et al, 1996).

Related business & engineering collaboratories

Automotive engineering design. We have done a series of studies at Ford Motor Company (funded by them) looking at distributed engineering design between North America and Europe. The strategy has been to do in-depth longitudinal studies of selected teams as they adopted distance technologies to facilitate their work (Olson & Teasley, 1996). We have also had an opportunity to study co-located teams collaborating in war rooms for the length of a project. These latter teams had huge productivity gains (Covi, Olson & Rocco, 1998.), setting a comparison point for other work arrangements. In a related project funded by NIST, we looked at distributed engineers in an auto parts company who worked in North America, Europe, and Australia. NIST will use our results to generate standards for engineering collaboration technologies.

Distributed software engineering. We began a five-year project at Lucent Technologies in 1998 looking at distributed software design among participants in North America, Europe, and Asia. Practical guidelines for using the technology effectively have been developed. An especially important issue, given the cultural differences, has been the formation of trust among dispersed collaborators (Rocco et al., 1999).

Global financial teams. We had an opportunity to study the use of Lotus Notes Team Room by global financial teams at IBM. A number of social barriers revealed how difficult it is to change work practices if the technology neither fits the culture nor has benefits directly for principal users. In the face of these obstacles, the work was reorganized to fit the geography, with difficult tasks assigned to co-located teams (Olson & Olson, in press).

Record keeping in collaborative work environments. This is a study funded by the National Archives that began in 1998. It looks at five cases of collaborative work in order to analyze the systems, practices, and individual roles for capturing, organizing, and reusing documentation of collaborative work.

The Broader Experience Base

We are not alone in the development, deployment, and evaluation of collaboratories. A number of other collaboratories exist in a variety of fields (see reviews in Wulf, 1993; Kouzes, Myers & Wulf, 1996; Finholt & Olson, 1997; Bly, 1998; Bainbridge, 1999). This experience base is now ripe for analysis to extract broader principles and formulate design methods to guide future collaboratory efforts. In Table 1 we show some representative projects to illustrate the range of scientific communities that have attempted collaboratories. This again is only a sample, and in this case the list favors those funded by US agencies. There are a number of European collaboratory projects as well – indeed, two recent conferences in Europe (Vienna, Dec. 1999; Paderborn, Mar. 2000) have focused on these. No syntheses of this body of experience have been undertaken. Our goal is to do the syntheses and generate the principles that promise to make future development and deployment of collaboratories more effective.

Table 1. Some Representative Collaboratory Projects

Science Domain	Principal Funding (if known)	Reference
Physical oceanography		Hesse, Sproull, Kiesler, & Walsh, 1993
Worm community	NSF	Star & Ruhleder, 1994
Spectro-microscopy	DOE	Agarawal, Sachs & Jonston, 1998
Environmental molecular science	DOE	Kouzes, Myers & Wulf, 1996
Fusion	DOE	McHarg, Casper, Davis & Greenwood, 1999
Diesel combustion	DOE	Pancerella, Rahn & Yang, 1999
Argonne National Lab	DOE	Churchill & Bly, 1999
Materials microcharacterization	DOE	Zaluzec, 1997
K-12 Education	NSF	Edelson, Pea & Gomez, 1996
Medical informatics	NIH	Shortliffe et al, 1998
Microscopic digital anatomy	NSF	Young et al, 1996
Telemedicine	DOD	Kilman & Forslund, 1997

Specific Research Plans

We now turn to more detailed descriptions of the activities we will use to accomplish our goals. Our activities will include:

(1) Detailed comparative analysis of collaboratory projects through invitational workshops that bring together collaboratory researchers from around the world,
(2) development of a Collaboratory Knowledge Base of technical and social data and detailed findings from existing collaboratory projects,
(3) iterative development of general principles and design methods, with broad community participation,
(4) development of a software components approach to collaboratory development that would be easy to configure by end users
(5) launching new collaboratory testbeds that would help us test our emerging principles
(6) training of domain scientists as collaboratory developers through a certificate program in collaboratory building.

We will have several partners in this project. First, in order to achieve the vision of having collaboratories engage a broad set of scientists and to encourage underrepresented groups to enter science careers, we will work cooperatively with colleagues from Howard University on all phases of the project. To ensure that the collaboratory concept generalizes beyond traditional science to include business and education, we will work closely with Bell Labs of Lucent Technologies. They are already funding an ongoing project (see above), in which their researchers are working collaboratively with Michigan counterparts. Finally, to assist the promulgation of the collaboratory idea to the broad world of non-profits and community organizations, we will work cooperatively with the Alliance for Community Technology, funded by the W.K. Kellogg Foundation.

1. Detailed Comparative Analysis of Collaboratory Projects

Collaboratory research to date has consisted of experimental engineering, in which systems are built and evaluated. This is fundamentally a process of learning-by-doing, which inevitably incurs substantial risks of failure and high development costs. This is a natural way to begin a field, but the time has come to move to more principled development. Since the collaboratory experience to date shows that the fundamental idea behind collaboratories remains very promising and worthy of continued development, it is time to summarize and synthesize what we know to abstract principles to guide further work in the field

The studies to date vary enormously in how much and what kinds of data were collected about the technical development process, the actual experiences of users, and the ultimate success of the collaboratory in accelerating scientific progress. Data about both technical and social aspects of performance are necessary in order to develop broad principles. On the development front, many initial projects developed custom software, but others are now using commercial applications. On the behavioral front, our collaboratory projects we have made extensive longitudinal observations of the behavior of our participants both prior to the deployment of collaboratory technology as well as after the deployment of a succession of new tools. In some cases (e.g., SPARC/UARC), these observations span a number of years. The published literature on collaboratories provides a scattered and incomplete record of what actually happened in other projects. We expect that in many cases others who have carried out collaboratory projects have much richer insights into what happened than is available in the published literature. In order to collect this rich data, we need a framework for asking the right questions about usage, and a venue in which to collect and codify the findings.

The best way to extract broader insights from both our own data as well as the experiences of others is to hold a series of focused workshops that ask a systematic set of questions across all of these projects. The initial series of workshops will focus on analyzing the experience of collaboratory developers and formulating a broad framework for gathering the results together. We will conduct structured interviews with those involved in different collaboratories, using a preliminary framework that includes issues at several levels:

- the underlying network and distributed computing infrastructure,
- he applications developed and used,
- the modules of software functionality that turned out to be the most useful,
- the usage and satisfaction data from users,
- and the social and organizational issues that arose.

We expect that initial discussions may lead individual investigators to return to their data for further analyses that they report in later workshops. We would also expect the workshops to lead to the identification of other kinds of projects that need to be done to further refine the framework. Data collected in this phase will seed the Collaboratory Knowledge Base, described in the next section. Subsequent annual workshops will continue this integration.

We will draw the participants for these workshops from several sources. Obviously, to support our focal integrating responsibility in this effort, the Michigan researchers will be centrally involved. Outside invitees will come from several sources:

- the science collaboratories of the type listed in Table 1 above;
- researchers and developers who have attempted similar projects in corporations (like our Lucent partners), and
- those who have attempted to use collaborative technologies for distance learning, particularly in the sciences (like ourselves and our Howard University partners).

In addition, we wish to use the workshops as a forum for increasing the diversity of collaboratory builders and users. Therefore, participants will be invited from institutions who have been absent from earlier collaboratory projects, such as predominantly minority colleges and universities. Students and faculty at these schools stand to be important beneficiaries of collaboratories in terms of improved access to unique scientific resources and to research activity surrounding these resources.

To illustrate the kinds of issues that will be discussed in the workshops, we analyzed our own collaboratory experience for preliminary empirical generalizations and associated research issues. We would expect such lists to serve as seeds for the workshop discussions. Table 2 presents a small sample of such generalizations.

Table 2. Illustrative Empirical Generalizations

Generalization	Implications	Associated Research Questions
Successful collaboratories depend on social readiness: Willingness to collaborate Openness to change	The introduction of collaborative technologies often fails because social readiness is lacking	What social or organizational interventions could facilitate the emergence of readiness?
Successful collaboratories also depend on technical readiness Appropriate technology experience Adequate infrastructure	Individual and organizational experience with technology is critical for success	Are there forms of technical training that could allow for easier or cost-effective adoption of new tools?
Work that is tightly coupled (high dependence on fine-grained interpersonal communication, tasks that are ambiguous) is not possible to do at a distance today	Work needs to be designed to fit the geography of the participants	What aspects of tightly coupled work would inform the design of new tools that would allow tightly coupled long distance work in the future?
Collaboratories develop along a life cycle from experimentation, to prototyping, to introduction and adopting, to supporting collaboration in a production mode	Collaboratories are most vulnerable when they transition from one stage to another	How can awareness of life cycle stages reduce the risk of failure? What design methods and tools can accelerate the development life cycle?
The addition of participants, data sources, and support for coordination to collaborative environments tends to yield complex user interfaces	Collaborative environments can be daunting, particularly to newcomers	How can the complexity be reduced through effective use of perceptual and cognitive principles at the interface level and organizational interventions at the community level?
Collaborative technologies tend to be strong on the support of formal, scheduled interactions and weak on the support of informal, spontaneous interactions of the kind that happen when participants are co-located	Relationships at long distance lack elements that are critical for forming coherent, trusting organizations	While many attempts have been made to support awareness of others and more informal interactions, these efforts have not gotten beyond the laboratory; so the research question is why?
Lightweight fluid motion between interaction styles and modalities is important	If the transitions between styles or modes of work are not smooth, people will stop using the technology.	What standards, protocols, APIs, wrappers, and infrastructure frameworks are necessary for embedding and coordinating components?

2. The Collaboratory Knowledge Base

As just described, during the first stage of the project (Years 1 and 2), we will organize and analyze data from existing collaboratories. The sources of such data include

- grant proposals and reports on collaboratories,
- existing evaluation studies,
- data on the development process, and
- information that emerges from our workshops
- published material on collaboratories.

In some cases, collaboratories have large amounts of raw data about use. For those that don't, we hope that the discussions at these workshops will provide stimulus for more data collection. We are also interested in the process of collaboratory development. We have already gathered extensive data on the development process in UARC/SPARC by evaluating existing records and interviewing nineteen people who had various roles in UARC/SPARC since its inception. We expect to be able to gather more of this type material as well.

The knowledge base of existing collaboratory experience will serve two purposes. First, it will support our abstraction of common principles and critical success factors in the development of collaboratories, described in section 3. Second, it will serve as a research vehicle in how to design such knowledge bases in general. For example, our research on UARC/SPARC suggests that a great deal of information is captured during the course of collaboratory development (e.g. plans, working documents, notes from meetings, e-mail, and much unstructured data). These materials help to explain the development and decision making processes, but this information is dispersed and not readily available for extracting lessons about collaboratory development. It is a very heterogeneous collection of data, reports and readings--some of which is written explicitly for the base, and some of which are mere collections of naturally occurring artifacts created in the conduct of work. The challenge is defining the content, organization and access mechanisms for this material.

We will test methods such as automatic summarization of documents (McKeown & Radev, 1995; Radev & McKeown, 1998) and use of metadata to make sense out of the vast quantities of unstructured data that collaboratories generate. Text summarization is the process of identifying salient concepts in text narrative, conceptualizing the relationships that exist among them and the generation of concise representations of the original documents that preserve the gist of their content. Radev (Radev & McKeown,1998, Radev & al 2000) has extended this to multi-document summarization through automatic identification of cross-document rhetorical relationships such as source agreement, contradiction, elaboration, change of perspective, and development. We will build techniques and tools for automatic summarization of clusters of related documents in the collaboratory knowledge base, including background research papers, meeting minutes, and reports of ongoing research work. As a result, we will improve access to the knowledge base by both current collaboratory members and potential new participants who join the collaboratory while it is in progress.

This material will be placed in an on-line data repository or archive, to be known as the Collaboratory Knowledge Base. It will support our work and will be available for use by other researchers studying collaboratories. The project leaders will design the form of this repository as well as the process for populating it prior to holding any of our workshops. Populating the knowledge base will be a learning opportunity for graduate students who will assist in gathering this material and facilitating use by workshop participants. It will also be a resource for the participants in our summer certificate program (see below).

During the second phase of the project (years 3-5) we will monitor and analyze on-going use of the knowledge base. We will assess patterns of use as well as determine what information is important to capture in order to foster on-going learning. Similarly, as we evolve our empirical generalizations, framework, and hypotheses, this will produce summary material as well as analyses of the raw data and links to emerging case material.

3. Development of Principles

A primary goal in this project is the abstraction and codification of principles, heuristics, and design methods to guide the rapid deployment of successful collaboratories. These principles will span both the technical and the social aspects of collaboratories, ranging from communication protocols, middleware services, and user interface paradigms, to changes in work flow practices necessary to fully exploit the new medium. In particular, we seek both technologies and principles of design or customization that will reduce the dependence of scientific collaborators on specialized collaboratory developers for designing, building, and deploying such tools.

There are instructive historical analogs to our goal. Sieworek, Bell, and Newell (1982) published the book *Computer Structures: Principles and Examples*, which made a seminal contribution to the advancement of the understanding and practice of computer architecture. The first chapters of this book offered a conceptual and notational framework for describing the both the structure and behavior/function of computers at various levels of abstraction. The rest of the book contains more than three dozen case studies of machine architectures including canonical descriptions of the machine in the framework proposed in the first chapters. The book was a widely used resource for both teaching and research and laid the framework for much more quantitative and theoretical approaches to computer architecture now best exemplified 20 years later by the benchmark book *Computer Architecture: A Quantitative Approach* by Hennessy and Patterson (1996).

In the world of collaboratories, we are at a similar evolutionary stage. We will work with data and empirical generalizations about what has happened in the decade-plus of experience with collaboratories. These data are more qualitative in nature, and not as formal and formalizable as the computer architecture material. Nonetheless, we feel that we are ready to undertake the kind of systemization and conceptual integration needed to move the field of collaboration forward.

First, we intend to identify the key dimensions along which collaboration tasks and user characteristics vary. This will involve a detailed evaluation of the existing studies to pin-point those dimensions that provide the greatest predictability for effective and ineffective collaboration. For example, we have already identified that users that share common ground have a greater chance to engage in successful collaboration compared to those who do not (Olson and Olson, in press). Next, we will operationalize these dimensions to systematically explore how they interact with characteristics of tools and tasks. This will enable a fine-grained understanding of collaboration tasks. For example, Table **3** shows eight characteristics that contribute to the building of common ground and which communication media support them (Clark and Brennan, 1991). The table shows that face-to-face interactions allow for many ways to achieve common ground including visibility, which enables participants in a meeting to have a rich set of cues to know the state of another person's involvement. In contrast, the telephone and video conferencing media do not allow for such cues.

Having systematically explored such dimensions, we will analyze the results of existing studies to identify how users have, or have not been able to exploit the capabilities of various collaboration media, and overcome their limitations. For example, telephone users have acquired strategies to overcome the limitation of lacking visibility inherent in the phone medium. They use strategies such as explicit turn-taking, frequent oral feedback such as grunts (versus just nodding), and saying "hello" to solicit feedback when there are long pauses at the other end. However, the acquisition of such strategies is largely a trial-and-error process that can be time consuming. In contrast, we intend to analyze the nature of tasks using various media and explicate the general principles of behaviors that enable effective collaboration processes for different collaboration tasks through different media. These general principles will enable new users to rapidly learn how to be effective in the use of various collaboration tools, in addition to guiding the design of new functionality.

Table 3. Eight characteristics that contribute to achieving common ground and which communication media support them. (adapted from Clark and Brennan, 1991)

Medium	***Copresence***	***Visibility***	***Audibility***	***Cotemporality***	***Simultaneity***	***Sequentiality***	***Reviewability***	***Revisability***
Face-to-face	•	•	•	•	•	•		
Telephone			•	•	•	•		
Video Conference		•	•	•	•	•		
Two-way Chat				•	•	•	•	•
Answering Machine			•				•	
E-mail							•	•
Letter							•	•

4. Development of a Software Components Approach to Collaboratory Development

So far, building collaboratories has required considerable investment in architecture development, custom programming, and system configuration. These are the things that have made collaboratory development a handcrafted process, even when generic application packages and commercial off-the-shelf (COTS) components are used.

We seek to change this. In order to allow users to configure collaboratories to suit their own needs, we need software components with widely accepted APIs that allow these components to be assembled into configurations of useful functionality. A related strategy would be to build wrappers as a way of integrating separate applications into a useful suite of capabilities. For either of these strategies to work, the elements need to be robust, standardized, and readily available. We do not have sufficient prognostication ability to know whether these kinds of components will arise in the commercial marketplace, or whether they might more naturally appear in the open source software world. We definitely know that we do not have the wherewithal to create complete, vertically integrated commercial quality systems ourselves. Rather, we can use our workshop process to explore the kinds of functionality required in collaborative tools and surface the technical issues required to deliver such functionality in a robust and modularized fashion.

We have considerable experience with both building collaborative tools and using COTS packages. In the UARC/SPARC project we have built three distinct generations of tools, moving from stand-alone custom applications to Java-based web clients to light clients with substantial servers. We have built systems that employed a component architecture at all these levels, and more, from the OO-based "Collaborative Tools Workbench" supporting real-time interaction, to dynamic network routing protocols and servers, to Java-based plug-ins for browser-based data analysis systems, to Lotus Notes Domino component systems providing web-based distance learning systems. Similarly, we have been carrying out a sustained evaluation of emerging products. A preliminary inventory of commercially available components can be found at www.communitytechnology.org/databases/groupware/.

Our goal is to use our experience as well as that of other collaboratory builders to assemble a base level of collaborative functionality that can be user-configured to meet the needs of the particular task at hand and the users' level of experience. The system needs to provide for easy integration of COTS functionality, developed in other academic-based efforts or in the commercial sector. We are currently using our experience in building an instructional environment for UM classes to sketch out a Worktools environment that will provide an integrated collaboration environment. We are integrating a Lotus Notes Domino back end with a variety of capabilities, including commercial tools like Microsoft NetMeeting and Webex. We will initially test this in the SPARC project, and our experience will provide us with preliminary specifications for the kind of components we need. These initial findings will help us shape our software components strategy.

One set of barriers to adoption of tools that cut across institutional, and country, boundaries is the level of local support required for server systems. This entry condition can lock out potential users, even whole communities. We will develop a set of components that would allow for a wide range of local participation, from zero responsibility for server support, where the local user community only has responsibility for browser and plug-in support, to high levels of local participation, with concomitant increases in local performance, for instance through caching, local data analysis and visualization, and increased local network utilization. This allows us to think of collaborative systems as a set of web-based services, and raises the 'component' issue at a higher level, that of the correct placement on the network of various services. Experience leads us to believe that both issues are crucial for successful collaborative systems.

Thus, we will seek to identify the reusable components for collaboratories, focusing on the functions that keep getting reimplemented in separate applications or in custom collaboratory software. We will define an architecture for how these components might be integrated, and develop documented standards for their interoperability. In fact, in order to allow for varying degrees of customization, this architecture would be multi-layered. One outcome might be “wizards” that would facilitate the establishment, development, and maintenance of collaboratories through a series of user friendly interactive panels.

5. New Collaboratory Testbeds

Existing collaboratory projects offer a broad experience base from which to begin the process of analysis and synthesis of principles. But they have spanned a range of user communities and situations that is incomplete. We intend to launch at least two collaboratory projects with new communities in order to enrich the experience base and accelerate the extraction of principles. These projects will collect systematic data about network behavior, application performance, and user behavior.

The collaboratory projects at Michigan, for which we have the most complete set of observations, have focused on small teams, mostly interacting in real time, carrying out planned, focused tasks. Our new projects would explicitly seek to expand our experience. Specifications for which kinds of situations are most deserving of investigation will come from our workshops and emerging framework, described earlier. But some characteristics of situations that we already feel are likely to be important include:

Scale. Collaborations in science can involve any number of participants, from a few to hundreds of collaborators. One of the best examples of large-scale collaborations is high energy physics, where technical papers may have a hundred or more co-authors. The emerging ATLAS project at CERN has upwards of 1500 collaborators. Collaborations on this scale are usually subdivided into subprojects that ultimately may have focused groups of one or two dozen participants. But there are issues of coordination across the larger projects that require different kinds of support. A second example of a large-scale collaboration is open source software development (e.g., Raymond, 1999). A collaborative enterprise such as Linux has had hundreds if not thousands of contributors.

Rapid response. In the collaboratories we have examined so far the projects are typically well defined and planned in advance. While we expected in the SPARC project that we might occasionally see science assembled on the fly in response to special events like solar flares, in actual practice these have not occurred. Collaboratories ought to be ideal for science constructed on the fly in response to either man-made or natural events on the earth (e.g., a natural disaster), or in space. We are exploring the introduction of collaboratory tools to the earthquake community, those interested in the global environment, and social scientists studying social responses to startling public events.

Short hold-time collaboration. We believe that scientific collaboration increasingly requires researchers from different fields to join forces very rapidly and for a limited period of time to accomplish specific objectives. Support for such collaborations requires not only the appropriate technical infrastructure to allow rapid learning of use protocols, but the mechanisms required to establish what has been called "swift trust." (Brown and Duguid, 2000). Swift trust is the ability to create rapid and effective shared work arrangements when the stakes are very high, as in highly competitive business or research situations. The ability to establish swift trust has been noted as one of the most important factors in the success of the Silicon Valley region, which like the field of science, works on distributed and often shifting networks of alliances

Mobility. All collaboratory projects have focused on either the desktop or the laptop as the entry point to a collaboratory. These relatively full-function computing platforms allow considerable latitude to the developer in what kinds of capabilities to include. But recently we developed access to SPARC sessions from a small device that combined a cell phone with a PDA. Mobile access could be useful when monitoring an ongoing situation over a period of days or weeks, or when doing remote field work. We may explore these issues first in some prototype development and evaluation in house, then deploy new mobile systems in some representative field setting.

Semi-structured data. All scientific research depends on data, and the various fields of science have elaborate traditions or even formal standards for how data is structured, stored, and retrieved. However, there are less formal kinds of data that are of potential importance to both the practice of science and to science education. For instance, being able to capture scientific sessions and then replay them either for further scientific reflection or for training young scientists can be done with collaboratory software. But it is still difficult to index, abstract, browse, and retrieve such sessions. We plan to offer these capabilities to some of the test sites to follow out the technical and behavioral consequences.

Diverse participant access. One of the promises of collaboratory technology is to broaden the access to science by a number of different constituencies. Early collaboratory projects have focused on engaging science elites in order to ensure broad acceptance of the new tools. We feel the time is ripe to proactively go after the vision of broader participation.

(1) Allowing remote junior scientists to participate in ongoing scientific activity; there are of course complex issues here involving the social networks in science, and the incentives for elite scientists to interact with emerging scientists beyond their immediate social networks. Lave and Wenger (1991) pointed out that legitimate peripheral participation is an important aspect of learning. Collaboratories enable a wider circle of such peripheral participation by those learning the science.
(2) Including the participation of scientists from underrepresented communities in the U.S. and in developing countries. This has always been held out as one of the promises of collaboratories, but to realize it we must take proactive steps. To this end, we have formed a partnership with Howard University. We expect faculty at Howard to be directly involved in all phases of the work described here. We will also take advantage of the pre-existing relationships at both Howard and Michigan in South Africa to engage relevant scientists in South Africa in both existing and new collaboratories. We have seed funding for this from the UM-Kellogg Foundation Alliance for Community Technology (ACT), and will seek additional support from the Kellogg and other foundations to enlarge this overseas engagement.
(3) Providing access to science by the general public; access to things like real-time scientific activity or captured sessions of remarkable scientific events can stimulate interest in and understanding of science; we have explored some of this through an outreach and educational site in our SPARC project (Finholt, 2000).
(4) Extending collaboratory tools for a broader range of community activities; we will partner with the Alliance for Community Technology to find ways to make the knowledge and tools developed in this project available to both geographic and virtual community groups with strong needs to collaborate. This will make it possible to enlarge the impact of our work far beyond the science communities that are the focus of this project.
(5) Providing an environment in which students and scholars are encouraged to attend lectures and seminars in different research areas: e.g., computer scientists attending seminars in the life sciences, statisticians participating in social science seminars.

6. Certificate Program in Collaboratory Building

In order to facilitate the dissemination of our principles and practices, and encourage a wider range of scientific communities to become self-sufficient collaboratory builders, we will conduct a collaboratory builder certificate program at the School of Information. The goal is to give certificate holders sufficient knowledge and skills about the technical and social elements of collaboratory construction. This curriculum content will come from our work in all of the other activities of this project as well as from the normal evolution of the academic programs at the School of Information.

Here is our preliminary idea about how the certificate program will run. We will refine it based on our workshops and our initial trials.

The certificate program will have three phases. In phase one, participants will spend two weeks in Ann Arbor, starting in mid June. They will receive intensive training from our faculty in the technical and social aspects of collaboratories. They will acquire hands-on experience with the technologies and the methods for behavioral analysis and evaluation. We will teach them how to assess and then facilitate collaboration readiness. They will become familiar with existing collaboratories, through both our multiple collaboratory projects and those of others that are readily available. In phase two they will be asked to conduct a preliminary needs analysis and technical scoping of a potential collaboratory in their field of work. This will include interviews and observations, leading to case material about present practice as well as scenarios about how collaboratories might work. They will learn how to map these needs onto an appropriate mix of social and technical solutions. We will provide interaction and coordination during this period using representative collaboration technologies. In phase three they will return to Ann Arbor for a week at the end of August to report back on their own experience, listen to the reports of others, and participate in evaluation and feedback of these reports. Full participation in both sessions in Ann Arbor, a written report on their needs and scoping analysis, and an oral presentation of their work during the wrap-up session in Ann Arbor, are all required to earn the certificate.

The primary targets for the certificate program are post-candidate doctoral students, post-doctoral fellows, or junior faculty/scientists in any field of science. But we will consider other categories of science applicants as well. We will advertise the program each winter, with an application deadline of February 15. The faculty on this project will constitute the selection committee, and we will select participants based on their promise as propagators of collaboratories in their field. Students will pay tuition, and the Rackham School of Graduate Studies at Michigan will provide institutional sanction for the program. We will offer financial support to 10 students. This support will include tuition, travel & lodging for two trips to Ann Arbor, and a travel grant of $2500 to support the required fieldwork. At the discretion of the participating faculty partial awards may be given to more students. Others are welcome to participate at their own expense if they meet the qualifications. We will guarantee our partner institution, Howard University, one funded spot in each summer's program.

Mapping our Expertise onto the Research Areas

The Senior Personnel on this project have considerable relevant expertise related to this project. Table 4 shows how their expertise maps on to the six areas of research just described.

Partnerships

Howard University is the leading historically black college/university in the US. Several faculty from Howard will be fully involved in all aspects of this project under a subcontract to this grant. Howard researchers will be involved in the analysis of current, ongoing collaboratories. Scientists at Howard will engage in the development and evaluation of at least one of the new collaboratories. Some candidates are existing collaborations between Howard, Michigan, and Michigan State in the Great Lakes Mid-Atlantic Center for Hazardous Substance Research (which does not yet have the kind of rich collaboratory environment like those described in this proposal), and science education projects involving Michigan, Howard, and several universities in South Africa. Finally, we will ensure that at least one Howard participant attends each summer's certificate program, in an effort to build human capital in the area of collaboratory development among their researchers. The principal participants at Howard are:

John Trimble (PI), Assistant Professor of Systems and Computer Science
James Johnson, Dean of the College of Engineering, Architecture and Computer Sciences, and Professor of Civil Engineering
Muhammadou Kah, Professor of Information Systems and Analysis

Table 4. The expertise provided by the senior personnel at Michigan

Participant	***1. Detailed Comparative Analysis of Collaboratory Projects***	***3. The Collaboratory Knowledge Base***	***3. Development of Principles***	***4. Development of a Software Components Approach to Collaboratory Development***	***5. New Collaboratory Testbeds***	***6. Certificate Program in Collaboratory Building***
Atkins	●		●	O	●	●
Bhavnani	●		●	O		O
Cogburn	●		●		●	O
Cohen	●		●			
Finholt	●	O	●	O	●	O
Furnas	O		●	●		
Hardin	●		●	●	●	●
Hedstrom	●	●	●			
Jahanian	●		●	●		●
King	●		●			O
Olson, G.	●	O	●	O	●	●
Olson, J.	●	O	●		●	O
Radev	O	●	●	●		O
Resnick	O		●	●		
Teasley	●		●	O	●	O
Yakel	O	●	●			

● = primary resource
O = secondary resource

Bell Labs at Lucent Technologies has an active program of research on collaborative technologies, principally because of their substantial internal needs as a result of global product design and production. They are already funding a project at Michigan (see above), and are Lucent researchers are fully participating in that project. Lucent researchers will actively participate in the several phases of this work, and they offer the prospect of additional testbed opportunities both internally and with external clients. Dr. James Herbsleb will be our principal technical contact.

The **Alliance for Community Technology (ACT)** is a new strategic partnership between the W. K. Kellogg Foundation and the University of Michigan. The goal of ACT is to link social investors, academia, and community-based organizations to innovate in the application of information and communication technology (ICT) to improve the condition of people and their communities. With start-up funding of $6M the Alliance is now exploring issues such as the role of collaboration technology in enhancing and building social capital; the role of open source software and application service provider structure to provide more effective ICT infrastructure for under-resourced organizations; international learning collaboratories to serve development in southern Africa, and a digital library federation for the 30 Native American tribal colleges in the US. Partnership with ACT offers the opportunity to explore collaboratory design and application in the context of access to education including broader participation in science, as well as economic, and community development.

Table 5. Rough Time Line for Activities

Task	Year 1	Year 2	Year 3	Year 4	Year 5
Data compilation & community building	Several workshops	Annual workshop	Annual workshop	Annual workshop	Annual workshop
Collaboratory Knowledge Base	Framework	Construction	Update and maintain	Update and maintain	Update and maintain
Emerging principles	Preliminary formulation	Dissemination & revision	Iterate with experience	Iterate with experience	Iterate with experience
Collaboratory components	Use workshops to extract components	Formulate specifications for components; conduct in-house testing	Work with partners to create components	Deploy components	Evaluate deployment
Certificate program	Develop and prototype	Summer courses	Summer courses	Summer courses	Summer courses
First new collaboratory	Identify characteristics	Baseline studies	Deploy and evaluate	Hand-off	
Second new collaboratory		Identify characteristics	Baseline studies	Deploy and evaluate	Hand-off
Third new collaboratory (self-sufficient)		Identify candidate	Baseline studies; train participants	Monitor the assembly and deployment	Evaluate & hand-off

Results from Prior NSF Support

UARC/SPARC (NSF Cooperative Agreement IRI 9216848, Sept. 1992 to Aug. 1998, $4,455,329; NSF Grant ATM 9873025, Oct. 1998 to Sept. 2001, $2,440,000) [Atkins, Finholt, Hardin, Jahanian, G. Olson, others] Since 1992 these two projects have developed, deployed, and evaluated a collaboratory for research in upper atmospheric physics. In the UARC phase of the project the focus was on providing tools for real-time data acquisition "campaigns", in which scientists had access to scores of remote data sources plus could chat with each other about what they were seeing. Over time the technology evolved from a custom-built application to a Java-based custom viewer to what we call a "thin client", meaning a client that only uses a web browser with HTML in order to improve performance. Hundreds of data sources are available, including supercomputer models. For those with more powerful technologies 3-D and VR visualizations have been developed. The most recent development is the support of electronic workshops, in which participants discuss research findings, publications, the planning of experiments – in short, all the other aspects of science. This has required creating access to data archives and digital library resources. An outreach aspect of the project has been very successful in bringing upper atmospheric physics to schools and homes. Further information about the project is available at www.si.umich.edu/SPARC/.

Technology Supported Group Processes. NSF Grant IRI-9320543, "Understanding the group processes in technology supported group work." 1994-1999. $750,000. [G. Olson, J. Olson]. This work investigated a number of ways in which small groups are affected by various kinds of synchronous information technology. We investigated work with shared editors and electronic whiteboards, compared to standard whiteboards and paper and pencil. We investigated a number of situations in which the tasks varied (design, instruction, persuasion, negotiation), groups were either established or strangers, and either from the same culture or different cultures. In some of the studies, the groups were co-located, working in the same room; in others they were remote, connected by either high quality audio or audio with high quality video. The results showed that groups who worked with a shared editor produced higher quality work than groups using standard whiteboards, paper and pencil. Established groups working remotely through high quality video and audio perform as well as face-to-face, and those communicating by audio-only suffer only marginally. However, when the groups are made up of strangers speaking English as a second language, video significantly helps communication over audio-only. Negotiation tasks produce more individualistic (rather than cooperative) behavior when people communicate over text-based chat than when they meet face-to-face. Initial meetings face-to-face engender enough trust to counter this effect. This grant generated 14 publications in peer-reviewed journals, conference proceedings, and book chapters. Eight graduate students have been involved in this work, resulting in 3 dissertations completed to date. Approximately 15 undergraduate and Masters level students have also served as research assistants on this work. Many of the publications that came out of this line of research are reviewed in Olson & Olson (1999, in press).

University of Michigan Digital Library. NSF Grant xxx, from xxx to xxx, $xxxx [Atkins, others] This project explored an approach to creating a large digital library federation based upon a network of distributed intelligent agents carrying out policies based on economic principles. The collections and services in the library federation were assumed to be both geographically and administratively distributed as well as heterogeneous multimedia. The project created theoretical results, including auction structures, which are now in use in information-based e-commerce. The project created and evolved a large testbed that focused on providing resources to support inquiry-based science education in middle and high school. Versions of this testbed were enhanced with additional funding from the Kellogg Foundation and the NSF and have been deployed in public schools in both Detroit and Chicago. The project also helped build collections, insight, and software which is now used in the extensive University of Michigan digital library production services and which is helping shape the National Digital Library Federation of major academic libraries. The project also indirectly lead to the JSTOR Project which is now providing large archival digital collections of scholarly journals to academic libraries around the world. More details about the project, including a list of 65 publications, are available at www.si.umich.edu/UMDL/.

Medical Collaboratory. NSF grant ECS-94-22701, from March 1, 1995 through August 31, 1999 in the amount of $1,320,000, entitled " A Collaboration Testbed in Medical Image-Based Examination, Diagnosis, and Treatment". [Finholt & others] The project produced prototype web-based collaboration tools for remote, asynchronous review of medical diagnostic images from MR, ultrasound, and plain film modalities. Descriptions of these tools are found at www.si.umich.edu/medcollab/index.html. In addition, the project conducted examinations of conventional radiological practices and of new practices made possible by the prototype collaboration tools and by other commercial applications (e.g., digital radiology systems). Efforts supported under this grant, in addition to production of the collaboration tools, resulted in over a dozen refereed conference and journal publications, including papers presented at CSCW 94, ECSCW 95, and CSCW 96. Finally, two students trained under this grant have subsequently gone on to prestigious positions in academia and industry (Elizabeth Yakel, University of Michigan; Nelson Manohar, IBM T.J. Watson Research Labs). Further, Yakel's dissertation received the best dissertation award at the winter 1998 meeting of the Association of Library and Information Science Educators.

References

Agarawal, D.A., Sachs, S.R., Jonston, W.E. (1998). The reality of collaboratories. Computer Physics Communications, 110(1-3), 134-141.

Bainbridge, W.S. (1999). International network for integrated social science. Social Science Computer Review, 17(4), 405-420.

Bly, S. (Ed.) (1998) Special section on collaboratories. *Interactions*, 5(3), 31-72.

Brown, J.S. and Duguid, P. (2000) Mysteries of the Region: Knowledge Dynamics in Silicon Valley. In Lee, C-S., et al., *Silicon Valley*, forthcoming.

Churchill, E.F. and Bly, S..(1999) Virtual environments at work: ongoing use of MUDs in the workplace. In Proceedings of the International Joint Conference on Work Activities Coordination and Collaboration, pp.99-108.

Covi, L. M., Olson, J. S., and Rocco, E. (1998) A room of your own: What do we learn about support of teamwork from assessing teams in dedicated project rooms? In N. Streitz, S. Konomi, and H. J. Burkhardt (Eds.) *Cooperative Buildings.* Amsterdam: Springer-Verlag. Pp. 53-65.

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

Finholt, T.A. (2000). The Collaboratory as classroom: Challenges and benefits of online participatory science education. In the panel session "Participatory science via the Internet" (S.D. Teasley, chair), International Conference of the Learning Sciences, Ann Arbor, MI, June 14-17, 2000.

Finholt, T. A., and Olson, G. M. (1997) From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science.* 8(1), 28-35.

Hennessy, J.L., & Patterson, D.A. (1996) *Computer architecture: A quantitative approach* (2^{nd}. Ed.). Morgan Kaufmann.

Hesse, B. W., Sproull, L.S., Kiesler, K., & Walsh, J.P. (1993). Returns to science: Computer networks in oceanography. *Communications of the ACM* 36(August):90 101.

Kilman, D.G., and Forslund, D.W. (1997). An international collaboratory based on virtual patient records. *Communications of the ACM*, 40(8), 110-117.

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

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

McKeown, K., & and Dragomir R. Radev, D.R. Generating Summaries of Multiple News Articles. In Proceedings of ACM-SIGIR'95 (Seattle, WA, July 1995).

National Research Council (1993). National collaboratories: Applying information technology for scientific research. Washington, DC: National Academy Press.

Olson, G.M., & Olson, J.S. (in press) Distance matters. *Human-Computer Interaction*.

Olson, J.S., & Olson, G.M. (1999) Computer supported cooperative work (Pp. 409-442). In F.T. Durso (Ed.), *Handbook of applied cognition*. New York: Wiley.

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

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. *Interactions*, 5(3), 48-55.

Pancerella, C.M., Rahn, L.A., and Yang, C.L. (1999). The Diesel Combustion Collaboratory: Combustion Researchers Collaborating over the Internet. In proceedings of SuperComputing'99 (Portland, Oregon, USA, November 1999).

Radev, D.R., & and Kathleen McKeown, K. Generating natural language summaries from multiple on-line sources. *Computational Linguistics*, 24 (3), pages 469-500, September 1998.

Raymond, E.S. (1999) *The cathedral and the bazaar*. Sebastopol, CA: O'Reilly.

Rocco, E., Finholt, T.A., Hofer, E., & Herbsleb, J (1999) Out of sight, short of trust. Paper submitted for publication.

Shortliffe, E.H., Patel, V.L., Cimino, J.J., Barnett, G.O., and Greenes, R.A. (1998). A study of collaboration among medical informatics research laboratories. *Artifical Intelligence in Medicine*, 12(2), 97-123.

Siewiorek, D.P., Bell, C.G., & Newell, A. (1982) *Computer structures: Principles and examples*. New York: McGraw-Hill.

Star, S. L & Ruhleder, K. (1994). Steps Towards an Ecology of Infrastructure: Complex Problems in Design and Access for Large-Scale Collaborative Systems. In CSCW '94, Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work,, (pp. 253-264). New York, NY: ACM Press.

Wulf, W.A. (1989) "The national collaboratory – a white paper." Appendix A in *Towards a National Collaboratory*, the unpublished report of an invitational workshop held at the Rockefeller University, March 17-18, 1989.

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

Yakel, E., Cote, S, Finholt, T., & Cohen, M. (1996) Medicine in the dark: Obtaining design requirements for a medical collaboratory from observation of radiologists at work. Short Paper presented at CSCW 96, Cambridge, Mass.

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). Interactive collaboratories for microscopy and microanalysis. In Proceedings of Multinational Conference on Electron Microscopy-97, Portoroz, Slovenia, Oct. 4-8, 1997.