GA–A22619

THE WORK OF SCIENTISTS AND THE BUILDING OF COLLABORATORIES

by

S. BLY, K.M. KEITH, and P.A. HENLINE

JUNE 1997

GA–A22619

THE WORK OF SCIENTISTS AND THE BUILDING OF COLLABORATORIES

by
S. BLY,* K.M. KEITH, and P.A. HENLINE

This is a preprint of a paper to be presented at the Group 97 International Conference on Supporting Group Work, November 16–19, 1997, Phoenix, Arizona, and to be printed in the *Proceedings.*

Work supported by
U.S. Department of Energy
Contract DE-AC03-76SF00098

*Sara Bly Consulting

GENERAL ATOMICS PROJECT 3228
JUNE 1997

The Work of Scientists and the Building of Collaboratories

S. Bly
Sara Bly Consulting
24511 NW Moreland
Hillsboro, OR 97124
sara_bly@acm.org

K.M. Keith
General Atomics
PO Box 85608
San Diego, CA 92186
keithkm@gav.gat.com

P.A. Henline
General Atomics
PO Box 85608
San Diego, CA 92186
henline@gav.gat.com

ABSTRACT

Interviews and observations at four sites which form the DOE Distributed Computing Experimental Environment provide a basis for understanding how distributed computing technologies can influence and support remote collaboration of scientific experiments. Scientists and technicians at these sites will form early user communities of collaboratories, blending laboratories and collaboration to make experimental facilities accessible from a distance.

Understanding the work practice of experimental scientists in the DCEE project provides input to the design of technologies for these collaboratories. The size and customization of equipment and the rapid iteration of experimental parameters challenge traditional groupware approaches.

INTRODUCTION

The United States Department of Energy (DOE) research community has undertaken the goal of creating collaborative laboratories, called *collaboratories* [1,2], by using computing technologies to support geographically distributed collaboration in scientific facilities. Collaboratory testbeds will provide remote access to expensive and hard-to-duplicate facilities, ranging from electron microscopes to a tokamak fusion research experiment. The goal is for scientists thousands of miles away to participate at the experiment site. The expected value of these geographically distributed environments includes substantially increased effectiveness in doing science, shared use of expensive experimental equipment, and an enabling capability for analytical and high-value production use by industry.

Work Activity Study

Four separate initiatives under the Distributed Computing Experimental Environment (DCEE) project will provide electronic means for collaborators worldwide to benefit from the research capabilities of US DOE sites. The four initiatives are located at General Atomics (GA) DIII–D Tokamak Facility, the Pacific Northwest National Laboratory (PNNL) Environmental Molecular Science Laboratory (EMSL), the Lawrence Berkeley National Laboratory (LBNL) Advanced Light Source, and the Argonne National Lab (ANL) LabSpace.

Each of these labs already has a wide range of geographically remote collaborators. Thus, the need for technologies to support such seems obvious. At the same time, the patterns of work in scientific laboratories involves years of tradition and closely held processes. Most technologies for collaborative work to date have involved small, relatively stable, working groups focused on shared computing applications or group meetings [e.g., 3–8]. Scientific experiments involve not only data sharing but also access to equipment, physical samples, and frequently changing participants.

Cultural evolution in creating collaboratories may be as critical as computing technology development. The introduction of collaboratory technologies will impact the work activity and the work activity will influence the design and use of the technologies in ways that are different from many groupware situations today. Understanding the work activity as well as the technology is critical to the long-term success of the these projects.

Laboratory Testbeds

Under the leadership of Lawrence Livermore National Laboratory, Oak Ridge National Laboratory, the Princeton Plasma Physics Laboratory, and General Atomics, scientists will conduct remote operations at GA's DIII–D tokamak. This project will demand real-time synchronization and exchange of data among multiple computer networks in experiments that involve as many as 40–50 scientists and technicians at a time. A major goal of the collaboratory is the presentation of sufficient auditory and visual information associated with the control room environment so that remote staff at multiple sites can be fully integrated into operations.

The PNNL testbed will be based on instrumentation being developed for the EMSL project, specifically, two unique nuclear magnetic resonance spectrometers, which are large, highly shared items, and some small instruments used by a limited number of researchers in molecular-beam reaction dynamics. Thus, the characteristics of two related yet distinct scientific cultures, working with two quite different kinds of machines, will be examined.

The University of Wisconsin-Milwaukee in collaboration with the Spectro-Microscopy Facility at the LBNL Advanced Light Source will provide remote access to three analytical tools that provide spatially resolved chemical information at length scales ranging (depending on the tool and the technique) from 1 micron down to atomic scale. The collaboration that uses these instruments is fairly large and geographically distributed, with investigators from nine institutions.

The two application testbeds at ANL involve state of the art instruments for doing the characterization of material on a neurotomic level and an international collaboration at CERN, the European high-energy physics center. The second testbed detectors and apparatus will not come on line for physics use until after the year 2000. The early planning phases already involve 100 institutions and several thousand people.

This paper provides a description of the scientists' work activity at these four emerging collaboratories and implications of that activity for the design of collaboratory technologies.

STUDY

As part of the collaboratory development, a field study is underway to document the ways in which scientists are working as a basis for understanding how distributed computing technologies can influence and support remote collaboration [9–11]. The field study is in two phases, the first to look at current scientific work activity and the second to look at initial deployment of collaboratory technologies and their uses. The goal of the first phase of the study is threefold:

- Provide documentation on the current work activity of scientists and technicians who are expected to form the user community of collaboratories. A characterization of the work activity will be used in designing and evaluating collaboratory technologies.
- Provide a means of validating with the users themselves the representation and summary of their work and concerns. Feedback from these users will be important in verifying the understanding of their work.
- Offer collaboratory designers and developers an external perspective on the way scientists and technicians talk about their work.

Over the past year, Bly visited each of the sites at least once, being presented with technology plans and demonstrations as well as interviewing a range of scientists who expect to use the collaboratory technologies. Fifteen scientists and technicians were interviewed, eleven at the laboratory sites and four who are already collaborating from a geographic distance. Data from these visits include notes, audio recordings and transcriptions of interviews, still photographs, sample documents, Web page information, and, in one case, videotapes of experiment operation.

OBSERVATIONS

The need for a distributed scientific laboratory is motivated by research and engineering which requires the use of scarce facilities (e.g., large electron microscopes, synchrotron light sources, various types of particle accelerators, etc.), by a single experiment being unique due to its scale (e.g., a fusion research experiment), or most commonly by the geographic distribution of collaborators and industrial partners at several different institutions. Although the instruments and experiments vary considerably at the studied facilities, four characteristics of the work arise in each case and have implications for the design and use of technologies to support remote collaboration. These are

- Expensive and hard-to-duplicate equipment for data collection.
- Rapidly iterating and changing experimental parameters.
- Multiple person, multiple specialties needed for carrying out experiments.
- Collaborators who are already geographically distributed.

Expensive and Hard-to-Duplicate Equipment

The expensive and unique nature of the equipment at all four sites means that the facilities are in high demand by scientists world-wide. The operation of the equipment is not well-known, and newcomers are regularly in and out of the facility. For the collaboratories, issues of training and the ability to make use of local expertise are critical. Figure 1 shows one experimental setup with customized hardware.

The sophistication of the experimental equipment requires that scientists understand the use of the instruments in order to prepare experiments appropriately and to operate the instruments effectively. A typical experiment consists of three primary phases: design, data collection, and analysis. At all of the sites, the design often includes building some of the equipment apparatus specifically for the experiment itself and/or preparing one or more samples to test. For the collaboratory, this means that distributed collaborations must support the planning phase of design, equipment modification or creations, and sample preparation. Usually the samples are prepared and characterized at the researcher's or industrial analyst's home laboratory

Figure 1: A beamline at the Advanced Light Source facility requires precise alignment and sample placement.

but require substantial knowledge of the experimental apparatus and effects.

As scientists say...

> "...it is a state of the art machine...the price is sophistication of the instrumentation, difficulty of running, expense of building the equipment, and the necessary expertise to know how to design an experiment, how to run the equipment, how to get the most out of it, and having a person that can also talk many languages...you need to talk material science, you need to talk electron optics, you need to talk computer science...."

> "The instruments that we use for doing the science are very large, very complicated and very expensive. They involve very large accelerator structures...the instruments we build to observe the collisions very often require design and construction efforts of from 7 to 15 years to build them to the point that you start doing the science..."

> "There are so many degrees of freedom here and what to explore efficiently takes experience to know."

> "Because we have so many different people coming through and every time anyone comes through, they go, 'Well, could we just change it a little bit?' OK, so as soon as you say that, you're inventing and you're into aluminum foil and you're into labor intensive experimental activities rather than sort of routine analysis type science that you might see in the lab of a semiconductor company where they have a machine for a specific purpose in a room and if it doesn't do what you want, why you go somewhere else. But these machines are sitting at the ends of the powerful x-ray beam lines. The machine doesn't do what he wants, you'll haul it out and put another machine in or change it because you need that facility, the beam line facility."

Rapidly Changing Experimental Parameters

Experiments at these four collaboratories are generally aimed at exploring an unknown space of physical phenomena made possible by the particular experimental apparatus. Unlike experiments in which many repetitions are required to obtain statistical validity, each successful run of an experiment yields significant results. Thus, each event in the experiment is usually followed by another with a change in parameters (or a fix of systems gone astray). This need to modify or change conditions rapidly and frequently has implications for the ways in which scientists can work remotely. Being closely in tune with the data as it's generated and with the control of the experimental conditions is critical.

As scientists say...

> "This is an instrument versus a machine...And I use that analogy a lot because a machine you put a sample in, you turn it on, you say Go do this, and then you go away. In an instrument, it's like a musical machine; you're playing it. And you're there all the time. You're doing a dynamic experiment. You're continuously changing the experiment. You're watching the data as it comes out. You're adjusting things, you're measuring things, you're changing the experiment as you go along."

> "The way it works obviously is that you don't think up everything beforehand and just kind of follow what you thought what would work, right? You make small improvements, or maybe even very substantial changes in the tactics as you go, right?"

> "But really, we're sort of reinventing it here...if this ever became a common technology, these tricks would get built into the product, to the hardware or software. Then anyone could run it. But right now they're not so you have to be in somebody's head. So there's quite a high level of skill required to drive it."

> "I don't know if there's any experiment that is that automated. Analysis is automated... There's no way to turn this into robotics. If it were possible to do robotics, it wouldn't be science anymore."

> "The data comes in very quickly here. In another beam line you might have a half-hour between operations on the experiment. You start something going, let it go half an hour while you have plenty of time to get a cup of coffee, get a print out, think about it, plot something up. Here, it's much more of a one minute time scale."

Multiple Persons, Multiple Specialties

The number of scientists and technicians involved in an experiment vary widely across the four facilities: from thousands of scientists at over 100 institutions to 40–50 scientists and technicians working on a fusion experiment (as in Figure 2) to a faculty member with post doctorates and graduates students using a microscope. Nevertheless, work at each of these facilities involves far more than a narrow understanding of physics or chemistry. Collaboratories will be made up of groups which include (a) the scientific lead team who designs the experiment and makes decisions minute-by-minute during data collection, (b) the operators who are responsible for ensuring that the equipment is working as specified, (c) the technicians who actually handle the equipment, and (d) the diagnostics teams who are responsible for the data collection.

Figure 2: The GA control room overview illustrates the size of the experiment team.

The lead scientific team is responsible for the overall experiment, monitoring events and making decisions from one run to the next to modify parameters or tune machinery to achieve their goals. The operators are responsible for implementation of the experiment, setting up the experimental parameters and ensuring the synchronization of the systems. The diagnostics teams are responsible for data collection, monitoring the sensoring equipment, ensuring data credibility, and measuring system outputs. Diagnostics teams may have experiments of their own which piggyback on the primary experiment. The technicians are responsible for the actual equipment, monitoring and fixing hardware.

> “So it's kind of like a NASA mission where everybody has their specialty and make sure that that part's working and then somebody-- in this case me--oversees putting all these together and making sure that everything is going to work together.”

A generalized view of the teams at General Atomics in Figure 3 suggests how 40–50 scientists and technicians may be actively involved in any experiment. In a situation more typical of the advanced light source or the electron microscopy experiments, two to five scientists take on all the functions necessary to run the experiment.

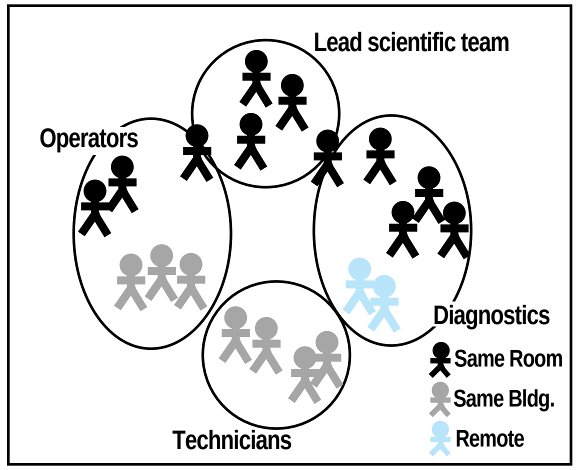

Figure 3: For experiments on equipment as large and complex as the magnetic fusion tokamak, many scientists and technicians are needed.

Collaboratories will have to support these diverse needs and communication paths. The degree to which various people must be closely in sync with one another during an experiment is a particular challenge for supporting remote interactions.

> “You're pretty much maxed out mentally to try to figure out what the hell's going on when you're doing these experiments. And what you really look for is someone who is running down thought processes in parallel with you, cross-checking what you're doing, thinking the same things that you are... so that's when it's really great, you work together with someone or three people and you're all up to speed and everybody's thinking and doing stuff together.”

> “We're responsible - very often individuals are responsible for some subsystem within a large detector. A large detector may have 50 subsystems. A group of physicists here might be responsible for one of those 50 subsystems. We designed it, built it, inserted it into the apparatus. We monitor it, maintain it, fix it when it breaks. We calibrate it.”

> “I usually participate in things that I have a scientific interest...Take responsibility for making sure the data's good, monitoring the data. Friday, even though I couldn't tell what was going on, I was...monitoring the...x-rays and making sure that they were running right and I could do that even not knowing what the experiment was....”

> “The bare minimum, two people have to be here. Bare minimum to get anything done. One person who knows how to run the experiment and one person who knows how to plot it.”

> “...where the person on the other end, if they have some specialized knowledge on how to interpret some data, then a lot of time we take data and we look at it and we need to have some

interpretation in order to continue on and know what we have."

"Some guys like to build things, huge big arrays of detector elements, and that's their contribution. Other people like to work on the data, work on visualization aspects and try to find ways of analyzing the data and so there's a fair amount of specialization in the field now and in fact one of the problems is just, you know, keeping up with the changes even within a collaboration not only within the whole science, but to be able to find out what's going on."

Geographically Distributed Collaborators

While the collaboratory work is designed to address issues of geographically remote collaborations, much of the current experimental work already involves scientists and technicians who may not be face-to-face during an experiment. At the General Atomics fusion laboratory, the scientist responsible for the day's experiment cannot be in the same room as the experiment, nor in the same room as many of the operators, technicians, and physicists running diagnostics (40–50 people). Compare Figure 4 with Figure 2, both from the same experimental run day. Even at the smaller end of the scale, scientists are often in offices away from the laboratory containing the experimental apparatus. In all cases, collaborators already exist throughout the world. The ways in which they are currently managing distance collaborations is a first step in building collaboratories.

Figure 4: For the fusion research experiment, people *and* equipment are distributed throughout the building.

"... I don't anticipate forever having to communicate with a walkie-talkie...but we still need to have an intimate relationship with the session leader and the physics operators who are planning the next shot...It's just a fast way to talk to them back there and let them know what's going on..."

"I sit over there sometimes so I can understand what they're talking about and even can anticipate what they might be thinking about ...so that kind of information that no one's going to sit down and bother to type on the screen, to take that much time."

[On working with a geographically remote colleague] "Yeah, it's not a very smooth communication, frankly. I mean, one of the reasons it hasn't been as successful as I had hoped is because our mode of communication is not very efficient...Usually we're working off a paper which is a, you know, a flat diagram, not the real structure — it's hard to communicate what we need."

"This morning we were having...computer problems. We only heard little tidbits of what was going on. But had we known, say, it was going to be down all morning, we'd be off doing other things. "

When she knows there's a problem, "then we call the power supply techs which are down at the different power supply areas to tell them there's a problem and what it is. That's the first they know of it unless they're sitting there, which lots of time they are. But all the techs know is that the breaker is open but not why or what happened so you definitely need to tell them what the problem is...."

[During an experiment, I go down] several times a day... I don't like a disembodied piece of information. I want to see it develop along....

While face-to-face interactions are the norm for communications during experiments, it is already not the case that all team members are copresent. The shadings in Figure 5 represent an example of distributed situations already occurring. Often operators and technicians are not in the same room as the primary scientific team; those running diagnostics may not even be in the same building. Figure 5 also suggests a few of the ways in which participants may be linked during an experiment. There's an opportunity for collaboratories not only to provide new remote collaborations but also to improve information flow in existing situations.

IMPLICATIONS FOR COLLABORATORY TECHNOLOGIES

The four DCEE project sites are each designing and implementing technologies to support distributed scientific facilities. Each of these projects will provide significant pieces of the systems needed to create long-term stable collaboratories. Furthermore, each of the projects represents different experimental sites and different technology needs.

Perhaps the most significant implication of the field study observations is the ways in which the collaboratory work differs from much of the work which collaboration technologies support today. Much of the CSCW research focuses on small on-going working groups, often in meeting situations.

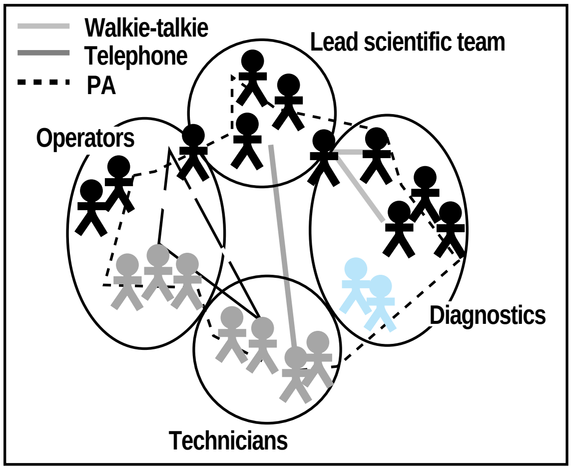

Figure 5: People work face-to-face or linked by telephone, walkie talkies, and the building PA.

The work activity is characterized by small groups which exist together on an ongoing basis, which typically interact in meeting situations (including discussions and presentations), and which do not rely on special shared equipment other than shared computing applications.

The collaboratory sites involve both small and very large groups which are constantly changing as new experiments come and go. Furthermore, the real-time interaction involves not only person-to-person communication but control of complex equipment and data in a variety of formats [12]. The aspects of the work at these experiment facilities suggest three challenges for the design of collaboratory technologies that are often not addressed in traditional groupware.

1. Resources are already in high demand. Because the equipment is often costly and unique, it is highly sought after. Furthermore, the training necessary to use the equipment requires local expertise to participate in all experiments and/or to provide training to visiting scientists. Providing ways for remote scientists to have greater access to equipment will have to be balanced with better ways to provide the local expertise needed to support the work.

2. Few have access to everything they need even now. Because the groups are often large and involve a wide range of scientific specialties, the flow of information is difficult to manage. Furthermore, data sets and visualizations of those data require formats and infrastructure that are hard to provide across a variety of computing platforms. Providing ways for remote scientists to access information suggests first ensuring adequate information exchange in existing collaborations.

3. Opening up collaboration implies exposing work-in-progress. In addition to issues of privacy and security discussed in CSCW technologies more generally, the experiment processes are frequently changing as they proceed. The context and history of the activity over time is critical to the understanding of the activity of the moment. Providing ways for remote scientists to participate in the work activity will involve finding ways to reveal information when and where it's appropriate.

The good news is that the technology and work practice can co-evolve. Remote collaborations already exist among the project scientists and these can be built upon. Suggestions arising from the work activity include small steps that can have immediate value:

— build on existing collaborations, enhance their abilities to work together

— improve intra-facility interactions.

— provide flexible data sharing before, during, and after experiments

— take advantage of simple solutions (i.e., transmitting the information meeting at the beginning of the day)

— establish critical mass with early adopters.

SUMMARY

The work described should have no surprises either for the scientists and technicians themselves or for he collaboratory development teams. However, it certainly has implications for collaboratory technologies, their evolution into use, and the resulting support of scientific experiments. The expensive and unique nature of the equipment, the complexity of the experimental design and setup, the range of specialties needed to conduct the science, and the rapidly changing experimental parameters all create a fragile balance of control over the experiment itself. A practice in which hands-on operation and finely tuned processes have dominated does not immediately lend itself to casual inspection and indirect operation. It is not enough for collaboratory technologies to make experimental facilities more accessible; the ways of working, the flow of control, and the positions of responsibility must also be taken into account and supported.

At the same time, there are already many geographically distributed collaborators and a need, as yet not well satisfied, for supporting shared analysis and results. The first steps of any collaboratory should be to support these collaborations and to build on the lessons learned from them.

The scientist who said

> "If you could open up a window on a workstation and actually see their laboratory in front of you or if they're using a workstation...if you could get control of their workstation, you could do the experiments yourself."

may not have thought through all the complexities surrounding these scientific laboratory experiments. The process of building collaboratories will be more than the development of distributed computing environments and more than the means for supporting distributed communications. The people, both those at a geographic distance and those working on site, are widely diverse in their needs for connection and their capacity to add overhead to the already difficult work of scientific research. Ongoing scientific collaborations will involve an evolution of not only the technology but of the work activity embracing that technology.

ACKNOWLEDGMENTS

We want to thank particularly the collaboratory teams at each site and the people who have provided introductions and background understanding: Bill McHarg, Deb Agarwal, Rick Stevens, and Jim Myers. We will thank the scientists and technicians for being willing to talk about their work and their environments. This is a report of work supported by DOE Contract DE-AC03-76SF00098.

REFERENCES

[1] National Research Council (1993): "National Collaboratories--Applying Information Technology for Scientific Research." National Academy Press, Washington, DC.

[2] Olson, G.M., Atkins, D., Clauer, R., Wegymouth, T., Prakash, A., Finholt, T., Jahanian, F., and Rasmussen, C. (1997): “The Upper Atmospheric Research Collaboratory (UARC).’’ In Coordination Theory and Collaboration Technology, G.M. Olson, J. Smith, and T. Malone (Eds.). Hillsdate, NJ: Lawrence Erlbaum Associates.

[3] Bly, S., Harrison, S., and Irwin, S. (1993): Media Spaces: ‘‘Bringing people together in a video, audio and computing environment.’’ Communications of the ACM, 36(1).

[4] Kraut, R.E., Egido, C., and Galegher, J. [1990]: “Patterns of Contact and Communication in Scientific Research Collaborations.” In Galegher, J., Kraut, R.E., and Egido, C. (Eds.), *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, Erlbaum.

[5] Moran, T.P., Chiu, P., Harrison, S., Kurtenbach, G., Minneman, S., and van Melle, W. [1996]: “Evolutionary Engagement in an Ongoing Collaborative Work Process: A Case Study.” In *Proceedings of CSCW ‘96*, Boston, MA.

[6] Olson, J.S., and Olson, G. [1995]: “What Mix of Video and Audio is Useful for Small Groups doing Remote Real-time Design Work?” In *Proceedings of CHI ‘95*, Denver, CO.

[7] Olson, J.S., Olson, G.M., Storrósten, M., and Carter, M. [1992]: “How a Group-Editor Changes the Character of a Design Meeting as well as its Outcome.” In *Proceedings of CSCW ‘92*, Toronto, Ontario.

[8] Tang, J.C., Isaacs, E.A., and Rua, M. [1994]: “Supporting Distributed Groups with a Montage of Lightweight Interactions.” In *Proceedings of CSCW ‘94*, Chapel Hill, NC.

[9] Blomberg, J., Giacomi, J., Mosher, A. and Swenton-Wall, P. (1993): “Ethnographic Field Methods and Their Relation to Design ” In Participatory Design: Principles and Practices, D. Schuler and A. Namioka (Eds.). Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 123-154.

[10] Holtzblatt, K. and Beyer, H. (1993): “Representing Work for the Purpose of Design.” In Representations of Work, L. Suchman (Ed.). HICSS Monograph (Hawaii International Conference on System Sciences).

[11] Plowman, L., Rogers, Y., and Ramage, M. (1995): “What are Workplace Studies For?" In *Proceedings of ECSCW'95*.

[12] Kraut, R.E., Miller, M.D., and Siegel, J. [1996]: “Collaboration in Performance of Physical Tasks: Effects on Outcomes and Communication.”. In *Proceedings of CSCW ‘96*, Boston, MA.