Cooperative Knowledge Work and Practices of Trust: Sharing Environmental Planning Data Sets

Nancy A. Van House, Mark H. Butler, Lisa R. Schiff
School of Information Management and Systems
102 South Hall #4600
University of California, Berkeley, CA 94720-4600
{vanhouse, markhb, lschiff} @ sims.berkeley.edu

ABSTRACT
Knowledge communities of all kinds have social and material practices for deciding what is known and who is to be trusted. In this paper, we address a specific kind of knowledge work, environmental planning, and a particular form of collaboration, the sharing of measurement data sets. We are interested in how trust is created; how trustability is assessed in the arm=s-length collaboration of sharing data sets; and how changes in technology interact with those practices of trust. We look at several elements of scientific practice that facilitate this sharing -- the publication system, communities of practice, boundary objects, and assemblages -- and discuss how a Web-based Digital Library might affect these elements and the knowledge work that they support.

Keywords
Knowledge work, environmental planning, boundary objects, assemblages, communities of practice, data sets, digital libraries, UC Berkeley Digital Library

INTRODUCTION
Changes in technology often foreground previously taken-for-granted practices. The World Wide Web is making possible new forms of cooperative work, and in so doing is causing us to re-examine some practices that make cooperation possible.

A major task of knowledge communities is deciding what is known, and what are the methods of deciding what is known, what counts as evidence, and who speaks. This, for example, is a major topic of science and technology studies (STS) [e.g. 3, 5, 15].

Knowledge communities of all kinds have both social and material practices for inscribing, transporting, evaluating, and using local knowledge, for bridging local and global, private and public, coordinating work across space and time. It is our contention that similar practices operate in many, perhaps all, areas where people are engaged in collaborative knowledge work, and so these processes can help us to understand some of the needs and concerns that arise in sharing information, particularly via the Web and similar networks that may be accessed by heterogeneous users.

Reference: CSCW '98: The ACM Conference On Computer Supported Cooperative Work. Seattle, WA, November 14-18, 1998, Proceedings; ACM, 1998, pp. 335-343

The primary focus of our work is knowledge work, which is social, distributed, and collaborative. Even the lone scholar is drawing on and contributing to others= work. However, much knowledge work is explicitly and overtly collaborative, involving people who may or may not know one another contributing to the joint construction of understanding in an area. We are particularly interested in the Aassemblages@ of people, practices, tools, theories, standards, social strategies, and the like that are dynamically, mutually constituted in and by knowledge work. We are concerned with how changes in technology ripple through these assemblages. And we are interested in how this understanding can help to shape technology that facilitates knowledge work.

One much-discussed aspect of the World Wide Web is a blurring of the boundaries between private and public, published and unpublished, closed and open. The Web makes it possible for people to post for public distribution information, documents, and data that would previously have been unpublished, private. It allows people to by-pass, or at least alter, the established channels by which communities have decided what to pay attention to. It opens up the collaborative process in ways that may be both intentional and unintentional.

In this paper, we address a specific kind of knowledge work, environmental planning, and a particular form of collaboration, the sharing of measurement data sets. Sharing data sets, as with other forms of collaboration,

requires trust. We are interested in how trust is created; how trustability is assessed in the arm=s-length collaboration of sharing data sets among people who may not know one another; and how changes in technology interact with those practices of trust. We look at several elements of scientific practice that facilitate this sharing -- the publication system, communities of practice, boundary objects, and assemblages -- and discuss how a Web-based Digital Library might affect these elements and the knowledge work that they support.

Much of cooperative technology and CSCW research has been concerned with relatively closed groups [12]; even when membership is fairly large, it is often nevertheless limited and/or known: e.g. people working on a project, within an organization, or as part of a professional community. Schmidt and Bannon [12] note that cooperative working groups are often large or embedded in larger organizations; transient; with membership that is not stable or sometimes even determinable; with dynamic patterns of interaction; distributed physically in time and space, and logically in terms of control; and with no omniscient agents organizing the work. They say that the Amajor thrust of current practical efforts to apply computer-based technology to cooperative work settings is directed at problems of supporting indirect and distributed cooperative work relationships, for instance ...in scientific research, where actors often cooperate at >arm=s length,= without direct communication and without necessarily knowing each other@ [12, p. 16].

In this paper, we explore some of the social practices of arm=s-length collaboration, among strangers, across space and time, foregrounded by the Web. We are specifically interested in the social and organizational practices of trust raised by the sharing of scientific data sets in environmental planning. We inquire about the practices that have facilitated data sharing in the past and how those practices do or don=t translate to the World Wide Web, and the implications for open collaborative technology like the Web.

THE PROJECT

The UC Berkeley Digital Library project is creating a digital library in support of environmental planning. Our goal is to develop technologies that support "work-centered" digital library services, oriented to the mission of the work group. Work groups require sophisticated digital library services and collaborative support to effectively utilize massive, distributed repositories of multimedia information. The technologies provide integrated, seamless access to bodies of new and old information via the network.

We are creating a digital library testbed of environmental information, consisting of a substantial and diverse collection of many kinds of textual documents, photo images, botanical data sets and various kinds of map data including aerial photos, digital ortho-photoquads, and digitized topographic maps. Our testbed currently holds 58,000 photo images plus 1,000 aerial photographs. We have botanical data sets of 8,300 recognized plants in California and 82,000 other plants worldwide. We have scanned and processed about 2,300 documents comprising about 251,000 pages of text.

We have made research progress in a the areas of image retrieval, new document models, natural language processing, and Web-based geographic information systems. We have constructed Web-accessible Java prototype systems of our research achievements to access our collections. The digital library, as well as added documentation about it, is publicly-accessible at http://www.elib.cs.berkeley.edu.

A major goal of the user needs assessment and evaluation part of the project has been to better understand the practices of environmental planning, particularly those of information use and production; the information artifacts used and produced; and the possible effects of a Web-based digital library (DL) on planning practices. More generally, our goal is to understand distributed, collaborative cognitive work, the role of information and information practices and artifacts, and the social context, uses, and nature of DLs. The work described in this paper is a part of our on-going research on knowledge work, environmental planning, and DLs.

METHOD

The focus of the research reported here is the work of a group of people engaged in environmental planning in California, particularly water and watershed planning. Our participants included professionals employed by state, local, and federal agencies, representatives of environmental organizations and industry, and concerned residents and landowners. This analysis is based, in particular, on discussions with (1) people engaged in two interagency (state, federal, and local) programs for sharing data sets; (2) state employees whose responsibilities included collecting, analyzing, and distributing data about various aspects of the environment; and (3) a group of local stakeholders, including local agency employees, landowners, residents, and representatives of environmental organizations engaged in planning for a specific and contentious watershed.

We conducted interviews in people=s own workplaces, observed meetings, and examined reports and other information artifacts used and produced by our respondents. Although we had also hoped to observe people going about their work, this was feasible only to a very limited extent. First, while some respondents= work consists primarily of collecting and analyzing data, for many others generally only a small part of their work

consists of visibly using or producing this kind of data. Furthermore, we found that many aspects of their work were politically sensitive. We were asked not to videotape public meetings, for example, because our informants feared that the presence of a video camera, however thoroughly we explained our purpose, would affect the already-contentious debate.

In this paper we only discuss the part of the study that had to do with sharing data sets. We did not begin by targeting this as an important concern for DLs. Rather, as we talked with people engaged in environmental planning about their work, their information production and use, and the possible uses of DLs, respondents immediately targeted data sets as an area of great concern where a DL could be useful.

ENVIRONMENTAL PLANNING

Environmental planning is often done by professionals employed by government agencies. In these settings it is often a collaborative, multi-disciplinary activity undertaken by people trained in various scientific and technical disciplines (e.g. engineering, biology) and in planning, a social science-based discipline. However, in California, as in other places, environmental planning often involves community-based groups representing the major stakeholders, including, in addition to government agencies, resource-based industries such as agriculture and timber, other environmentally-based industries such as recreation, residents, landowners, and non-government environmental groups. The goal is for a wide range of stakeholder groups to come to a limited shared understanding of the current state of their geographical area, the threats to it, and the likely outcomes of various future actions, and to reach a loose and perhaps shifting agreement on actions to be taken by the various parties.

The planning that we have studied generally does not consist of a single effort with a fixed group of participants and a defined beginning, end, and product. Instead, we find a dynamic, on-going, overlapping array of programs and projects, with participants recombining in various subsets, coming and going as the issues and personnel change. Planning issues and projects cross multiple levels, for example, state, regional, and local planning efforts interact in various ways. The process is often crisis-driven: such events as winter floods or a proposed listing of a fish species as endangered will mobilize action.

It is also explicitly, overtly political. The stakeholder groups have diverse and often-competing interests. Relations among the various actors may be delicate and strained. The political and economic stakes are high. As are the environmental stakes: many of these decisions have serious long-term environmental consequences that would be difficult or impossible to reverse.

The result of planning is often not a blueprint but a shared understanding. Participants often do not need to -- and generally will not -- come to an all-embracing agreement, only to a negotiated shared understanding, Acooperation without consensus@ [12]. In the watershed that we studied, for example, most of the land is privately-held. The likely actions are not so much regulatory as voluntary. The state of the watershed is a function of people=s day-in and day-out practices; a major goal of environmental planning is to ensure that participants engage in practices consonant with the health of the watershed. The shared understanding becomes invisibly embedded in people's interpretations and practices [4].

Environmental planning is collaborative, not just in the planning process itself, but in the creation and use of information. Environmental planning is characterized by uncertainty and disagreeement. The usual first step in watershed planning, for example, is to infer from limited evidence the current state of the watershed, its problems, threats, and trends. Possible causes and solutions are considered and disputed. A wealth of information and information artifacts are used and created: data sets, reports, maps, tables, memos, newsletters, and so on.

Information is used, not just for making decisions, but for defending points of view and persuading and educating others. It is an ally, in Latour=s terms [5,6]. It serves interests. It is contested. Different groups used and privileged different information and uses of it: engineers, for example, were somewhat contemptuous of the Asoft@ data of biologists, and both were suspicious of the social science approaches of professional planners.

In this paper, we focus on one kind of information that is used heavily in environmental planning: measurement data. Much of what we learned applies to other kinds of information, as well, but we will concentrate primarily on this kind of information, how it is used, the concerns expressed by our respondents, and the implications for collaborative work.

Examples of the kind of data that we are talking about include indicators of water quality, population samples of fish species, the distribution of plant species, and reports on water supply. Data are collected using measurement devices and/or human observers, often at the same locations over a period of time, with a fixed or variable frequency. Data collection is often expensive, time-consuming, and requires skill. The data go through a complex chain of transformation and representation from raw data to summarization, analysis, and interpretation, creating a cascade of representations [5]. These data are used to describe past and present, to find trends, to test hypotheses of causality, and to forecast the future.

Data are collected at varying levels of granularity, ranging from satellite photos to a person walking the ground and

mapping each individual plant representing an endangered species. The level of granularity often correlates with the organizational level at which data are collected: the federal government may be the source of satellite photos, local residents may who walk the ground and map individual elements. This means that cooperation across levels is necessary.

Long time series and data on multiple variables for the same site and/or period of time are particularly useful for tracking trends and assessing causal links. Analysts may combine data from a variety of sources, often collected by a different observers at different times using different methods and definitions. Combining data from multiple sources is difficult, requiring understanding the similarities and differences among multiple sources and how to appropriately combine them.

A current concern may result in people going back through old data sets looking for evidence, e.g., if a fish species seems to be declining, historical data on population and on possible correlating factors may suddenly become useful. As a result, a major concern among many of our participants was tracking down unpublished data from a variety of sources. A recurring assumption was that a wealth of potentially-useful data was buried in file cabinets around the state.

Until recently, data were collected and published primarily by government organizations and large non-profits. The unanalyzed data were available to only a few people, usually only directly from the owner of the data. Most users had to be satisfied with published summaries, analyses, and graphical representations. Without access to the underlying data set, users= abilities to re-interpret or re-use the data or challenge the analyses were limited. The data collected by other groups and individuals was held privately and shared only to a very limited extent.

Computing and telecommunications are changing the means by which data are distributed. In California, a variety of efforts are underway to identify and share environmental data sets. File transfer protocols have been used to share data among agencies. More recently some groups are experimenting with using the Web to distribute data to all comers.

CONCERNS ABOUT QUALITY CONTROL

Of the concerns raised about wide sharing of data sets, in this discussion we will focus on what many respondents loosely called Aquality control.@ These concerns were of two sorts: misuse of data, and the distribution of poor quality data.

Misuse of Data

Many professionals were concerned that making unanalyzed data accessible via our Web-based DL could result in inappropriate use of the data or incorrect interpretations, with potentially-serious consequences. (One respondent spoke of Ahaving our data used against us.@) For example, one source of data on fish populations is boat trips that net fish and measure their catch. The time of year and day, the stage in a species' spawning cycle, weather conditions, type of net, the depth at which the net is trolled, and many other factors affect the catch. A comparison with "last year at this time" and conclusions about trends in the fish population that did not account for these factors could be seriously misleading.

Currently, technical and organizational barriers to data acquisition help to ensure that unanalyzed data are primarily available (1) to Aqualified@ individuals, and/or (2) with appropriate explanation and advice. For example, most data are available only in summary form. To get the raw data one must contact the agency. This contact allows the data supplier to advise the user about the nature of the data, limits to its use, and so on. The difficulty reported with this system is the time and labor required of both supplier and user and the power that this gives the data supplier in controlling who uses the data.

Provision of "Bad" Data

Another major concern was the ease with which data collected by untrained individuals using inappropriate protocols or differing definitions of data elements and methods could be promulgated. Professionals worried about untrained users' inability to evaluate such data or the source. The concern was about not only poor quality data but inappropriate analyses and interpretations, e.g., combining data from multiple sources in inappropriate ways.

Others saw the democratization of the data supply as a necessary counter-measure to what they saw as the limits and biases of "official" sources, making finer-grained measurement data available and empowering citizens to take a more active role.

PRACTICES OF TRUST

To better understand these concerns, we looked to several sets of practices and social relations that, in the past, helped environmental planners to collaborate: the publishing system, and scientific practices.

The Publishing System and Libraries

One institution that has coordinated the work of people who are distant in space and time is what Chartier calls the Aorder of the book@ [1, 9] -- which includes the publishing system and libraries. It makes accessible the work of people who may be unknown to one another in a form that is transportable and reliable -- Latour=s Aimmutable mobiles.@ [5]. With measurement data, what is made available is most often summaries, syntheses, and graphic representations, not raw data B Latour=s inscriptions [6] Publication acts as a multi-stage system of filtering and reduction: first creating inscriptions, then

determining which will be published, then what will be made available and preserved in libraries. Each step is an exercise of cognitive authority, deciding what is worth sharing and preserving, and making linkages among data, documents, authors, and ideas. Whether we define this as quality control or social control, the publishing system mediates between producers and users of information. It enforces some kind of (varying, indeterminate, contestable) standards. Technical and material barriers to publication act as social barriers, as filters to ensure that information meets some group=s standards.

Many of the people we talked with in local agencies and interest groups were more interested in our Digital Library as a forum for publishing their materials than as a source for acquiring others.= Getting one=s own point of view heard is a major task in the contentious world of environmental planning.

Some environmental publications go through such channels as peer-reviewed journals. Much more is distributed in self-published reports. State and federal agencies and large, well-funded environmental organizations generally have more resources to distribute and gain attention for their data and analyses than do, say, local environmental groups or dissenting individuals. Libraries, in choosing which reports to make accessible, often give priority to Areputable@ organizations and government agencies.

Scientific Practices

Scientific practices are aimed at ensuring objectivity, the Aview from nowhere@ that establishes a fact as a fact, at closing the black box on how a statment became a fact [5, 7, 10]. These are practices of what counts as evidence, logic, and proof, who speaks, and whose judgment counts. Whether or not we believe that the view from nowhere is possible [3] is irrelevant for our discussion here; scientific practices are designed to translate local knowledge into global, personal into universal [17].

The Aorder of the book@ and scientific practices are intertwined. Latour [5, 7] goes so far as to describe publications as primary products of laboratories.

We have found several concepts that are current in the literature on CSCW, scientific practices, and situated learning useful in understanding the issues around Web-based access to environmental data sets: communities of practice, boundary objects, and assemblages. Together they give us a way to talk about the concerns of the professional environmental planning community around the increased openness of the Web.

Communities of Practice

A community of practice is a group (e.g., work group, profession, or discipline) that shares practices, understandings, technology, artifacts, and language [8]. Communities of practice are where professionals learn the methods, attitudes, values, and interpretations of their community. What is learned is not simply a body of knowledge but a mode of action, an ability to improvise appropriately . They learn not only how to practice their profession but how to be a professional and how to understand the world. Communities of practice are about, among other things, what is known, what counts as evidence, and who speaks; and how it is decided what is known, what counts, and who speaks. AA community of practice is an intrinsic condition for the existence of knowledge, not least because it provides the interpretive support necessary for making sense of its heritage@ [8 p. 98].

Their model is apprenticeship. The apprentice learns, not only from the master, but by doing, watching, listening, and from other members of the community, including other apprentices. The master too learns continually in this process. And not just the apprentices and the masters but also the skills and knowledge are changed in this process.

The community and its practices and tools are dynamic. Communities both reproduce and transform themselves by bringing in new members. There is a continual tension between stability and change, transformation and reproduction, old-timers and newcomers.

The boundaries of communities of practice are permeable. A person typically belongs to several: a botanist who works for the State doing research on a particular species in a particular geographical area may belong, with varying degrees of centrality, to several intersecting communities of practice. The interaction among communities of practice as they recombine around problem areas to create emergent, fluid new communities is part of the tension between change and stability.

Communities of practice are sources of trust. One way that users judge whether work or information is to be trusted is to look at its source: is he or she a part of our community of practice? Can he or she be trusted to have used accepted methods to collect, analyze, and interpret the data? Do we speak the same language? Do they see the world the same way that we do?

Environmental planning is distributed knowledge work engaged in by people from multiple communities of practice, separately and together. Part of the tension in environmental planning is the conflict among these. We see a major goal of a planning project or group as creating a (limited and temporary) community of practice among the people concerned with a particular geographical area and/or issue. The group, in working together and making collective sense of the evidence about their area and issue, forges a shared set of practices and understandings about

who they are, what they are doing, and about their area of common concern.

Boundary Objects

Collaboration takes place in environmental planning both within but across groups. Multiple groups from different disciplines, organizations, and geographical areas are concerned with, for example, the population of coho salmon in western US coastal waters, or the impact of agricultural run-off as it works its way downstream and out into the ocean.

Environmental planning is not unique in needing to reconcile multiple viewpoints and create solutions that are coherent across groups without necessarily reaching complete consensus. Star suggests that one tool used by scientists (and, we would argue, other knowledge communities) are what she calls Aboundary objects@ [13, 14]. Boundary objects are shared by multiple communities. They are both plastic enough to adapt to local needs and have different specific identities in different communities, and robust enough to maintain a common identity across sites, and thus to be a locus of shared work.

Star identifies a non-exhaustive list of four types of boundary objects, which are helpful in fleshing out the meaning of the term:

- Repositories, which are Aordered piles of objects indexed in standardized fashion@ [13, p. 48]. They are modular (that is, users can can lop off subsets), and the units of analysis are heterogeneous. Examples include libraries and museums.
- The ideal type or platonic object, which, being abstracted from all domains, doesn=t describe any one of them in detail but whose vagueness allows it to cross domains. Her example is early atlases of the brain, which in fact described no brain but communicated across the research and clinical worlds.
- Terrain with coincident boundaries and different internal contents; e.g., the state of California was a boundary object for the people who collected specimens, funded, managed, and did ecological research using the Museum of Vertebrate Zoology [14]. All the participants were interested in the one way or another in natural science within the boundaries of California.
- Forms and labels, such as standardized forms filled out in one site and passed on to another for further work.

Although some have interpreted Star=s presentation of the negotiated character of boundary objects as requiring that the users interact, we would argue that one strength of boundary objects is their ability to coordinate the work of people who do not interact directly.

We argue that data sets are boundary objects, and so understanding boundary objects will help us to understand the concerns about shared data sets. A set of observations transformed into a set of records using standardized descriptive categories, terminology, and classification have the characteristics of both a repository and of forms and labels. In environmental planning, people are continually using data sets collected by other people and organizations, often remote in place and time. Data sets have to be constructed so as to be useful both to their original owners and to known and as-yet unknown others.

In environmental planning, as well as in other areas where data are shared, a tremendous amount of attention is devoted to metadata. The issue is not simply to calibrate measurements, terminology, and data elements across data sets, which are important, but also to try to attach to each data set sufficient descriptive detail to allow the user to understand and interpret the data appropriately in order to engage in the delicate work of using multiple data sets in the same task, and the same data set in multiple tasks. Whether such detail can be sufficiently specified is, however, debatable.

Assemblages

Another approach to the question of how local knowledge gets Arendered equivalent, general, and cohesive@ in science and in other knowledge systems is that of Watson-Verran and Turnbull [17]. They define an assemblage as Athe amalgam of places, bodies, voices, skills, practices, technical devices, theories, social strategies, and collective work that together constitute technoscientific knowledge/practices.@ They say that no other term quite gets at this concept -- that epistemes, paradigms, actor networks, and boundary objects all address some part of assemblages. They find the term useful in that it connotes:

> ...the ad hoc contingency of a collage in its capacity to embrace a wide variety of incompatible components. It also has the virtue of connoting active and evolving practices rather than a passive and static structure. It implies a constructed robustness without a fully interpreted and agreed-upon theoretical framework....Assemblages constitute connections and contrive equivalences between locales in knowledge systems....Here the relations of power and knowledge are understood as invested in the material, social, and literary practices of discourse and representation, discipline and resistance. [17, p. 117]

Their examples of assemblages include those of Gothic cathedral builders (who designed and built large, complex, innovative structures without architects or plans over

hundreds of years), the Incas (who coordinate the work of and fed 5 million people in an arid environment without reading and writing), and the knowledge system of Pacific navigators that was likewise embedded in an oral culture. All are presented as examples of Athe melding of quite heterogeneous and disparate practices to form stable assemblages that connect@ [17 p. 126], Adisciplining life at the local level in constituting a knowledge system.@ In other words, assemblages are collages of disparate elements that coordinate local actions in such a way that the local knowledge becomes global.

What is useful for our purposes is the heterogeneous, *ad hoc*, dynamic nature of assemblages, and how they coordinate people=s actions such that local knowledge is transported, shared, coordinated, and connected. Assemblages include people, processes, practices, tools, information artifacts, theories, and other elements that work together in determining what is known, and the relation between local knowledge and that which is accepted more broadly.

SHARING DATA SETS

To share data sets is to grant them the status of boundary objects. It is to determine that they are sufficiently robust to be used across sites, and malleable enough to be adapted to local needs. It is to understand them enough to know when and how to use them.

No amount of metadata, narrative text, or even personal contact between data suppliers and users can fully document what a user needs to know to completely understand a data set and trust its producer. Instead, this understanding and trust is rooted in the community of practice and the assemblage of environmental planning. The originator is expected to provide enough information to allow the user to determine that she or he can trust the originator and therefore the data. Part of the grounds for that trust is knowing that the originator is a part of one=s (or of a reputable) community of practice. When professionals expressed concern to us about data collected by citizens, they didn=t know that the collectors failed to follow accepted protocols, but they couldn=t be sure that they did. It=s not simply a matter of following instructions: Goodwin (1994), for example, shows how archaeology students on a dig are taught to Asee@ what an archaeologist sees in the dirt. The seeing is learned through situated practices in relevant settings.

This trust is also rooted in the assemblage of environmental planning, the network of practices, understandings, tools, and so forth. When both the creators and the users are grounded in this assemblage, both know, for example, how a data set fit into this assemblage: that it is a set of sample observations that follow certain standards of sampling and data collection; the sample design is based on certain assumptions and theories about the natural work; the data are operated upon and represented in standard ways; they are used as inputs to models; and so forth. To say that the data need to be understood in context, while true, understates the complexity and richness of the assemblage that provides that context.

PUTTING DATA ON THE WEB

Now we can look again at the concerns expressed about provision of poor quality data, and inappropriate use of data by means of a Web-based Digital Library. It is our contention that the concerns that we heard expressed represent concerns about the divorce from the publishing system, the increased permeability of the boundaries of communities of practice, the dissociation of data sets from the assemblages within which they are created and are intended to be used, and the questioning of their status as boundary objects.

The Web can (thought doesn=t necessarily do so) by-pass many of the social and technical processes by which communities decide what is known, what is to be trusted, what is accepted as Apublic,@ published information. Because it creates different technical barriers to sharing information, it also changes the social barriers. People outside the community of practice -- or whose status is unclear -- can distribute their information more easily, and can access information intended for members of the community within an assemblage of practices and artifacts.

Schmidt and Bannon [12] raise three key issues for cooperative work involving a large and indefinite number of people, from different perspectives, over a protracted period of time: identifying the originator of information, as a means of assessing its validity; identifying its context of production and use; and identifying the politics of data: data, they say, are not innocent and neutral. Our environmental planners would heartily agree.

These are all important questions. It is our purpose to show some of the richness and complexity of these questions, and that they are of course not unique to CSCW. Rather, CSCW highlights and questions some of the embedded practices by which people have answered these questions.

The problem addressed here has implications for other kinds of collaborative knowledge work and systems. For example:

- Sharing of databases of many sorts, not just of measurement data, is increasingly common, within and across organizations, professional communities, and the like.

- Many organizations are now engaged in what they call knowledge management. The term is used in many ways, but it often involves giving

people access to reports and other artifacts created in other parts of the organization for other clients and projects, at other places and times -- in essence opening people=s file cabinets. When is it appropriate to do this, and when is the loss of context and of the nuanced understanding of the originator dangerous?

ACKNOWLEDGEMENTS

This work was supported in part by the NSF/NASA/ARPA Digital Library Initiative under contract # IRI-9411334, Robert Wilensky, Principal Investigator.

REFERENCES

1. Chartier, R. *The order of books: readers, authors, and libraries in Europe between the fourteenth and eighteenth centuries.* Trans. by L. Cochrane. Stanford University Press, 1994.
2. Goodwin, C. Professional Vision. American Anthropologist, 96:3, 1994, 606-633.
3. Haraway, D. J. Modest_Witness@Second_Millenium. ©FemaleMan_Meets_Oncomouse™: Feminism and Technoscience. NY: Routledge, 1997, 23-39.
4. Innes, J. Introduction to the Transaction Edition: Understanding the Alliance of Knowledge and Policy. In: *Knowledge and Public Policy: the Search for Meaningful Indications*, 2nd expanded edition, New Brunswick, NJ, Transaction Books, 1990.
5. Latour, B. *Science in Action*, Cambridge, MA, Harvard University Press, 1987.
6. Latour, B. "Visualization and Cognition: Thinking with Eyes and Hands", in *Knowledge and Society: Studies in Sociology of Culture Past and Present*, 6, 1986, 1-40.
7. Latour, B. and Woolgar, S. *Laboratory life: the construction of scientific facts*, intro by Salk, J., Princeton, NJ, Princeton University Press, 1991.
8. Lave, J. and Wenger, J. *Situated Learning: Legitimate Peripheral Participation*, Cambridge, England, Cambridge University Press, 1991.
9. Levy, D.M. "Cataloguing in the Digital Order" in *Proceedings of Digital Libraries '95*, Austin, TX: Hypermedia Research Laboratory, Texas A&M University, 1995.
10. Potter, T. "Information, power, and the view from nowhere", in Bud-Frierman, L. editor, *Information Acumen: the understanding and use of knowledge in modern business*, London and NY, Routledge, 1994, 217-230.
11. Schiff, L., Van House, N. and Butler, M. "Understanding Complex Information Environments: a Social Analysis of Watershed Planning." *Digital Libraries '97: Proceedings of the ACM Digital Libraries Conference*, Philadelphia, PA, July 23-26, 161-186, 1997.
12. Schmidt, Kjell and Liam Bannon. "Taking CSCW Seriously: Supporting Articulation Work", *Computer Supported Cooperative Work*, 1:7-40, 1992.
13. Star, S. "The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving", in Gasser, L. and Huhns, M., etc. *Distributed Artificial Intelligence* 2, Pitman Publishing, 37-54, 1989.
14. Star, S. and Griesemer, J. "Institutional ecology, translations, and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39", *Social Studies of Science, 1,* 387-420, 1989.
15. Star, S. ed. *Ecologies of Knowledge: World and Politics in Science and Technology*, State University of NY Press, 1995.
16. Suchman, L. "Centers of Coordination: a case and some themes", to be published in Resnick, L., ed. *Discourse, Tools and Reasoning*, forthcoming.
17. Watson-Verran, H. and Turnbull, D. "Science and Other Indigenous Knowledge Systems," in Jasanoff, S. et al, eds., *Handbook of Science and Technology Studies*, Sage, 115-139, 1995.