Building a geocollaboratory: Supporting Human–Environment Regional Observatory (HERO) collaborative science activities

Alan M. MacEachren *, William Pike, Chaoqing Yu, Isaac Brewer, Mark Gahegan, Stephen D. Weaver, Brent Yarnal

GeoVISTA Center, Department of Geography, 302 Walker, Penn State University, University Park, PA 16802, United States

Received 22 February 2005; accepted in revised form 2 November 2005

Abstract

Collaboratories have been defined as centers without walls, virtual places where teams of scientists can undertake coordinated research. As part of the Human–Environment Regional Observatory (HERO) infrastructure project, we have been developing a geocollaboratory to support work by geographically distributed scientists about geographic problems. Our specific focus is on science teams developing and applying protocols for long-term study of the local and regional scale human impacts of global environmental change. The HERO geocollaboratory includes web and other Internet-based tools to enable same-time and different-time (thus synchronous and asynchronous) different-place collaboration. Methods and tools have been developed to support (1) synchronous distributed meetings that include video links and shared visual display of geospatial information; (2) asynchronous perspective comparison and consensus building activities; and (3) long-term information sharing and knowledge development. This paper introduces the research effort, sketches the conceptual framework within which the geocollaboratory is being developed, outlines progress thus far in the three collaborative components listed above, and discusses our experiences using these tools for distributed science as well as our plans for continued development. We direct specific attention to three web-based, collaborative tools we have developed in support of components 2 and 3 above: an

* Corresponding author.

E-mail address: maceachren@psu.edu (A.M. MacEachren).

URL: http://www.geovista.psu.edu (A.M. MacEachren).

0198-9715/S - see front matter © 2005 Elsevier Ltd. All rights reserved.
doii:10.1016/j.compenvurbys.2005.10.005
e-Delphi tool (supporting sharing and comparing of expert opinions), a concept-mapping tool that supports building, sharing, and comparing concept relationship diagrams linked to formal ontologies, and a web portal (called Codex) that provides a personal workspace, mechanisms for forming groups and accessing group resources, and methods for encoding knowledge objects that include geographic referencing. © 2005 Elsevier Ltd. All rights reserved.

1. Introduction

Understanding the local and regional impacts of global environmental change requires systematic monitoring and analysis at multiple locations over extended periods. The Human–Environment Regional Observatory (HERO) infrastructure project is addressing this challenge by creating a model for, and infrastructure to support, a network of distributed data collection and research sites (observatories). The effort includes four observatories, located at Pennsylvania State University, Arizona State University, Clark University, and Kansas State University.

A primary goal behind the HERO network is to build the information resources to support long-term scientific research partnerships. For the data collection and analysis efforts of the HERO network to be successful, geospatial technology and methods that are developed and implemented must meet two goals. First, the technology and methods must support collaboration among scientists at different HEROs as they work together on common problems. Second, the technology and methods must facilitate context- and task-sensitive data encoding and storage, knowledge construction and management, and data/knowledge retrieval in support of distributed science communities working collaboratively with distributed resources. An initial problem that HERO researchers are addressing collaboratively is the development of robust protocols for collecting, integrating, and applying data to complex problems in human–environment interaction. These protocols encapsulate shared scientific knowledge of the HERO researchers and others upon whose past work they draw. The protocols must be general enough to support cross-site comparative investigation of the impacts of environmental change, while remaining flexible enough to address site-specific questions. The protocols must also be adaptable enough to address as yet unformulated questions by future scientists.

We are addressing these goals, in part, through development of the HERO geocollaboratory, a distributed research-support environment linking the four HEROs and colleagues at other collaborating agencies (e.g. USGS, NOAA) in a “center without walls” (Cerf et al., 1993). In this paper, we outline the conceptual basis for development of the HERO geocollaboratory and detail progress toward its implementation. The focus in this paper is specifically on methods and tools developed to support multisite science teams working on creation and application of research protocols for collecting and applying data to understanding and comparing specific local implications of global environmental change. We have taken an iterative approach to design and implementation of methods and tools, with early experiences guiding decisions to revise, extend, or replace initial implementations. The methods and tools have not been assessed through formal tests. Instead, we have acted as participant observers in a gradual, iterative process of changing how science is done by inserting new technologies into the work of distributed science teams, observing what
works and what does not, refining tools, and reinserting. We report here on the first three years of experiences with, and insights gained from, these observations.

We begin, in Section 2, by reviewing the development of the collaboratory concept in science generally, to provide the context for our work. Section 3 provides an overview of the HERO geocollaboratory, a description of specific collaboratory components implemented thus far, and reflections on use of these components by the HERO team. In the first part of this section, we discuss methods and tools for supporting both same-time (synchronous) and different-time (asynchronous) work as well as the HERO team’s initial experiences with the collaboratory tools and the use of those experiences to refine the tools. A key input to our method and tool development process has been the application of these methods and tools to support a distributed Research Experience for Undergraduates (REU) Site. Twelve students participate in REU activities each summer (three per HERO site). We focus in the second part of Section 3 on use of geocollaboratory tools during the first and second summer of REU student work. In the final section, we discuss planned extensions to the HERO geocollaboratory and outline a set of research questions that initial experiences have generated.

2. Background

The challenge of building national collaboratories was detailed in a 1993 National Research Council report. This report introduces the collaboratory concept by quoting an unpublished white paper by Wulf (1989) who defines a collaboratory as a “...center without walls, in which the nation’s researchers can perform research without regard to geographical location – interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information from digital libraries” (Cerf et al., 1993, p. 7). Considerable progress has been made toward the report goals through development of prototype collaboratories to support research in physical science (Keahey et al., 2002; Kouzes, Myers, & Wulf, 1996; Olson et al., 1998; Russell et al., 2001; Schissel et al., 2002), health science (Craver & Gold, 2002; Olson, Teasley, Bietz, & Cogburn, 2002), and computational science (Kaur, Mann, Matossian, Muralidhar, & Parashar, 2001).

A primary focus in initial collaboratory efforts has been on developing and implementing the networking technologies required to make remote connections possible and secure (Kaur et al., 2001). A common objective has been to enable remote connections to laboratory instruments (Henline, 1998) and to simulation models (Schissel et al., 2002). More generally, emphasis has been on shared technologies to facilitate real-time data collection or control of experiments in the physical or medical sciences.

Current technologies for different-place work vary in how they cope with distances in space and time. Asynchronous communications like e-mail and threaded discussions via mediated bulletin boards impose minimal constraints on participants’ schedules, but are not effective for achieving consensus in a short time. Recent developments, which focus on integrating visualization with text, show promise for asynchronous collaboration (Neuwirth, Morris, Regli, Chandhok, & Wenger, 1998). Synchronous methods require participants to assemble at appointed times. Instant messaging (e.g., ICQ) and chat rooms (e.g., Internet Relay Chat) allow only limited use of visual display (a key factor in discussion of geospatial data); and traditional video conferences require participants to be in a specially equipped room. Audio and video conferencing that makes use of the Multicast Backbone
(MBone) through the Internet remove these impediments (Rhyne & Fowler, 1998). Our early experience using MBone to facilitate collaboration in virtual environments suggested that collaborative sessions combining this technology with collaborative visualization tools can be very effective (MacEachren et al., 1999), but we also found that (at the time) not all US universities were able to support MBone connections.

Technologies for remote communication and collaboration have reached sufficient maturity that attention is beginning to be paid to understanding the influence of these technologies on how joint work is done (e.g., Gutwin & Greenberg, 2004). While much of the attention has been directed to group work in business activities (e.g., Mark, Carpenter, & Kobsa, 2003), the impact of technologies that enable remote collaboration on science work is also beginning to be considered (e.g., Fuhrmann & Pike, 2005; Watson, 2001). Early studies of the impact of e-mail on science work provide evidence that the increased interaction made possible by e-mail does enhance scientific productivity (Walsh & Roselle, 1999) but that it does not have an impact on the relative balance in productivity – those who are most productive continue to be similarly productive relative to other scientists (Finholt, 2002).

With some collaboratories moving beyond the experimental implementation stage to regular use, initial efforts are being directed toward exploring how collaboratories influence group work and whether specific collaboratory implementations have been successful and/or have involved a wider range of scientists in core science activities relative to traditional collaboration. In a meta-analysis of 10 years of published research on computer-supported cooperative work relevant to collaboratories, Olson and Olson (2000) identified key aspects of individual and group dynamics that can lead to success or failure of technology-mediated same- and different-place group work. Specifically, groups found to be most successful in taking advantage of collaboration technologies were those groups with an initially high level of common ground, that engage in loosely coupled work, and that have a readiness both for collaboration and collaborative technology. More recently, Sonnenwald and Li (2003) and Sonnenwald et al. (2003) explored aspects of social interaction among science students who collaboratively used a remote specialized scientific instrument called a nanoManipulator. They found that students with a strong competitive learning style were more positive about the advantage, compatibility, and complexity of the collaboratory system relative to same-place work. Students having an individualistic learning style, or working with another student having this style, were less positive.

In spite of the collaboratory developments detailed above, limited progress has been made toward applying or to understanding implications of the collaboratory concept in the social sciences generally or, more specifically, to the study of human–environment interaction (Kuhlman, Soffer, & Foresman, 1997). But, human vulnerability and responses to global environmental change are exactly the kinds of problem where collaboration across research disciplines and places is essential – thus a problem domain for which the collaboratory approach should be particularly applicable. One impediment has been a lack of work directed to fusing collaboratory concepts with developments in collaborative use of geographic information systems (Churcher & Churcher, 1999) or geovisualization (Rhyne & Fowler, 1998); see also MacEachren (2001) for more on map- and GIS-based collaboration. A second impediment has been a lack of a geographic/human–environment science research tradition in approaching research questions through large, collaborative science projects.
To address these issues, HERO geocollaboratory development and implementation has focused on: (a) the general issue of enabling perspective sharing and joint work among distributed human–environment researchers and on (b) the specific goal of enabling joint work to conceptualize and study human–environment processes. The primary human–environment science goal of the HERO project is to collaboratively develop a conceptualization of multiscale, human–environment interaction and the formal protocols for data collection and analysis methods needed to study the process in a way that supports multisite and longitudinal comparisons. Thus, emphasis in this paper is on application of the collaboratory approach to achieving these initial HERO project goals of problem conceptualization and research protocol development (rather than on using the collaboratory to study of human–environment processes).

3. HERO geocollaboratory components

The HERO geocollaboratory is being designed to support group work by scientists and future scientists in asynchronous activities that range from joint work by members of a research team on specific research tasks to long-term knowledge sharing across large science communities. Initial emphasis within our geocollaboratory development efforts has been on methods and tools to support three activities: (1) synchronous, distributed meetings that include video links among participants and shared visual display of geospatial information; (2) asynchronous perspective comparison and consensus building activities among both co-located and distributed researchers; and (3) long-term knowledge capture and sharing. Each is discussed below.

3.1. Synchronous distributed meetings

If all participants in a research project are co-located, regular and impromptu meetings play an important role in keeping everyone up to date, carrying out discussion of research ideas and methods, critically assessing those ideas and methods, and making important project decisions. To support distributed science teams, standard video conferencing technologies can meet part of the need. However, the nature of geocollaboration demands support for shared digital workspaces as well as real-time video links. Specifically, to address key human–environment issues concerning place and region, collaborating scientists need to share maps, images, and other graphic depictions of information relevant to discussions; and they need tools allowing them to annotate and manipulate representations presented by others (Brewer, MacEachren, Abdo, Gundrum, & Otto, 2000). To meet these needs, we have experimented with a range of technologies that support combined video and data connections using the Internet.

In this section, we discuss available network environment architectures and their general advantages and disadvantages for supporting the multipoint, simultaneous connections required for synchronous, distributed group work. Then, we detail our experiences with two kinds of network environment architectures (Unicast and Multicast, defined below) for application within the HERO effort.

3.1.1. Network environment architectures

Network technology plays a fundamental role in building a virtual collaborative environment for synchronous audiovisual and data conferences over the Internet. There are
three primary types of IP (Internet Protocol) network environments: Unicast, Multicast and Broadcast (Rouskas, 2000); also see www.cisco.com/warp/public/614/17.html.

An IP Unicast network is a point-to-point communication in which each piece of information is sent from one point to another single destination point within an IP-routed network. In Unicast environments, multimedia communications can be achieved through the H.323 protocols (i.e. a set of algorithms) that provide a foundation for audio, video, and data communications, for details see (the International Engineering Consortium www.iec.org/online/tutorials/h323/index.html). One important component defined by the H.323 recommendation is the multipoint control unit (MCU). An H.323 MCU provides the capability for three or more terminals and gateways to participate in a multipoint conference (see Fig. 1). Generally, the ability of multipoint communication in a Unicast network to support real-time video and data connections will be a direct function of bandwidth.

Multicast technology was originally motivated by Deering’s (1989, 1991) work on standards for multicasting. An IP Multicast environment provides dynamic one-to-many and many-to-many connectivity between various senders and a group of receivers, without consuming extra bandwidth for each (Fig. 2) (Almeroth, 2000; Cisco, 2002, chap. 43; Williamson, 2000). The Multicast Backbone (MBone) supports a collection of real-time and multimedia applications that can, in turn, support collaborative tools. These include: RAT – the Robust Audio Tool, which allows users to participate in audio conferences; VIC – a video conferencing tool; WB – a shared whiteboard tool; NTE – the Network Text Editor that is sharable on the MBone, either simultaneously or non-simultaneously; SDR – a session directory tool designed to allow advertising and joining particular sessions of MBone conference (Johnston, 1996; Savetz, Randall, & Lepage, 1996); see also http://dsd.lbl.gov/OldMisc/mbone/.

---

![Fig. 1](https://example.com/fig1.png)

Fig. 1. A centralized multipoint collaboration with an MCU (multipoint collaboration unit). With this setup, the MCU controls connections between user’s displays. The architecture is Unicast, with each node receiving a separate communication from the MCU.
IP Broadcast is a one-to-all technology that transmits the same packet of data to all nodes at the same time on a local area network (LAN). If some clients are not interested in the transmitted information, this technology will waste bandwidth and, if there are many broadcasts sent simultaneously across the network, a broadcast storm can result. A broadcast storm uses substantial network bandwidth and may cause network time-outs. Due to these drawbacks and the need to support many-to-many connections for HERO, we tested only Unicast and Multicast options.

3.1.2. Applications to HERO

We begin below with discussion of Unicast methods and tools. Unicast-based team meetings have become routine activities for HERO scientists. We reflect here on experiences covering three years of use. We have tested MBone-based Multicast tools, and report below on those tests. But Multicast tools are not yet used in practice by the HERO team because the Multicast network environment is not yet available at all four HERO sites.

3.1.2.1. Synchronous activities in a Unicast environment. As noted, current HERO synchronous collaborations are achieved in a Unicast network environment. Our earliest synchronous activities made use of a specialized video camera (Polycom ViaVideo), a Radvision hardware-based H.323 MCU, and Microsoft NetMeeting software for video and data conferences. This combination of tools was reasonably effective in supporting a range of situations from full-team video conferences, through small-group work sessions, to two-person impromptu information sharing sessions. More recently IBM Lotus SameTime, a software-based MCU, has been applied to our synchronous collaboratory activities. SameTime provides similar synchronous collaboratory functions to those in NetMeeting. This includes text messaging, streaming audio and video, a shared whiteboard and shared applications. We selected SameTime because it provided reliable connections and relatively
powerful data-sharing functions. However, because the version of SameTime we have used only allows the speaker’s video to be viewable, and its video screen size is limited, SameTime does not support a large group of collaborators effectively. We have found it to be most suitable for two-point collaborations via desktop computers. It can also support connections via mobile phones and wireless PDAs.

Below, we outline our experiences using synchronous collaboration tools. We divide discussion on the basis of meeting size (size of groups and number of sites) and formality (pre-planned versus impromptu). We have observed substantial variations among differently sized groups in the style of interaction, the success of specific tools, and the kinds of preparation needed for successful meetings.

Scheduled, large group meetings: To support HERO all-hands meetings, video conferencing tools have been particularly useful. Fig. 3 represents an all-hands, distributed meeting in summer 2002. After having conducted several such group meetings, we can suggest some guidelines for success. For a distributed group of this size, a tightly structured meet-

![Fig. 3: A synchronous collaboratory session – HERO all-hands, distributed meeting in summer 2002. Approximately 25 individuals participated in the meeting. Participants were located at all four HERO observatories. Additionally, one researcher participated by phone from the UK. Two nodes with live video and desktop computer access to the shared display space were active in Pennsylvania (PA), one each in Arizona, Kansas, and Massachusetts. As shown, the two PA groups were relatively large, with one or two individuals on line at the other locations. This particular meeting focused on discussion of the joint vulnerability analysis protocol being developed and experiences with its initial trial application in PA. Multiple individuals in PA presented their work to the distributed group. Both PowerPoint slides and web pages where shared and discussed.](image-url)
ing agenda is critical. During the meeting, the quality of audio and video is a key to achieving successful interaction among the sites; thus allotting time to test the technology prior to the meeting is important. Resource-sharing functions, to present PowerPoint slides or view a participant’s desktop, can help the meeting if pre-planned. Spontaneous resource-sharing remains difficult, because it requires both particular hardware and software configurations and some practice on the user’s part.

Sub-team, scheduled work sessions: In addition to supporting distributed, all-hands meetings, video conferencing and related tools were used to conduct scheduled, synchronous geocollaboratory activities among smaller groups (usually 3–6 people). Activities in the sessions included the exchange of ideas as research protocols were developed, general discussions pertaining to research presentations and publications, and discussion of specific research tasks such as the revision of interview scripts prior to field work. REU students were frequent users of the synchronous collaboration tools, scheduling weekly, Unicast video conferences to discuss how each person’s research was progressing. The collaboration tools (including support for sharing maps and other displays and using a combination of visual display, text, and voice discussion) allowed the REU students to negotiate a common understanding of vulnerability protocols that each group was attempting to apply to a vulnerability assessment of a local area and to use their different local experiences to suggest modifications to the protocols.

During video conferences, REU students recognized that their different skill levels, varied research interests, and diverse previous research experiences exacerbated the difficulties associated with distributed, collaborative research. The collaborative sessions helped students to understand the multiple geographical differences in their research sites (i.e., sites having different climates, ethnic backgrounds, economic characteristics, etc.). The students reported that they saved valuable time by using the video conferencing tools, vis-à-vis using e-mail or the telephone to accomplish similar goals. Several REU students discussed (during a video conference) that video conferencing software was the only suitable tool to negotiate how to incorporate individual methodological decisions and site-specific considerations in each research project while keeping the basic concepts and the meta-protocol for applying the concepts intact.

As with the full-team scheduled meetings, experiences over time with these smaller, scheduled collaboratory activities lead to some basic guidelines. Regular, synchronous, small-group activities are often most successful if complemented by other communication (e.g., e-mail) before and after the synchronous session. Before a meeting, e-mail communication was often used to clarify the tasks and to exchange necessary documents. Ongoing use of asynchronous collaboration tools (see discussions of e-Delphi and Codex below) helps provide context for synchronous exchanges. At the beginning of the meeting, it is useful for an organizer to list a clear agenda reminding participants about the primary purpose of the meeting and to suggest a sequence for addressing specific topics. At the end of a session, it is useful to summarize or reemphasize the major points. After the meeting, both e-mail and asynchronous geocollaboratory tools were useful to follow-up the meeting topics until goals were reached.

Spontaneous, small-group meetings: The REU activities generated many spontaneous video conferences to discuss time-critical issues or to have detailed discussions among 2–3 students. For example, students investigating the socioeconomic variability at each site – a key aspect of their research – used additional video conferences to complete the negotiations concerning modifications to methods so that the methods were both appropriate for
each geographic location and comparable among locations. The video conferencing software was a user-friendly tool that students enjoyed using because it was easily accessible on several computers at each HERO observatory and required little technical expertise. One REU student said that “the video conferencing software provides a quick and highly efficient tool to have discussions and build consensus.” Moreover, the students appreciated the personal contact with distributed team members gained through video conferencing; the software helped collaborators learn more about one another while its visual component limited the misunderstandings common to e-mail or even telephone communication.

For two-person collaboration, the meeting times and agenda can be very flexible. The focus of interaction is usually quite specific and data-sharing functions are often important to in-depth discussion and problem solving. Microsoft’s NetMeeting was used in conjunction with the video conferencing software to facilitate spontaneous discussion, resource sharing, and expertise sharing between REU students at different observatories. Although data were typically shared using more traditional media (e.g. e-mail or shared web repositories), NetMeeting provided a way to share computational resources and expertise synchronously. In one example, a student in Massachusetts requested the help of a student in Pennsylvania with a specific question regarding data analysis using GIS. The students initially tried using e-mail and the telephone, but the solution to this specific problem was too complicated to convey without a visual demonstration. The students then took advantage of the NetMeeting-supported video conference tools. The student in Massachusetts granted access to the student in Pennsylvania to gain control over her desktop. By discussing the steps verbally in conjunction with a visual demonstration, the necessary information was successfully conveyed remotely.

Two summers of experience with use of synchronous collaboration tools to support spontaneous, small-group meetings supports several observations. First, based on informal feedback over time by our participants (particularly by the REU students) and our own observation of selected meetings, the video and audio enabled collaboratory environment is often more successful than text-only chat (or e-mail) in achieving shared understanding or solving problems related to methods and technologies. This assessment, while not based on a controlled study, agrees with previous studies (e.g., Bos, Olson, Gergle, Olson, & Wright, 2002; Tscholl, McCarthy, & Scholl, 2005). Video–audio enabled tools seem to allow participants to clarify ideas and correct misunderstandings in a more natural way than through e-mail or chat and they provide better support for making conceptual connections across topics while making it less likely that the thread of an argument will be lost. In combination with desktop sharing, they also facilitated productive discussions of maps and mapping tools. Still, video-enabled collaboration does not equal face-to-face collaboration. In face-to-face situations, body language, mood and feelings are more easily communicated and impromptu follow-up is more practical. Current video–audio conferencing and application sharing tools impose substantial constraints on these implicit cues to meaning, especially when only the speaker’s video is viewable (as in SameTime).

3.1.2.2. MBone tools. Because the Multicast network environment has not been available at all HERO sites, the tests of MBone-based collaboratory tools were conducted locally. We tested these tools by simulating a use session focused on geospatial information analysis. The particular scenario tested was a joint analysis of some real spatial data about electricity production in the US using exploratory data analysis tools available in GeoVISTA Studio (www.geovista.psu.edu). We were able to confirm that the MBone-based support for
efficient video and data communication can enable joint science work with geospatial applications.

The Mbone-based collaboratory tool applied in our tests was CORE2000 (http://collaboratory.emsl.pnl.gov/download/core2000/core2000.html), which is a Java-based cross-platform environment developed by the Pacific Northwest National Lab. In addition to incorporating tools such as RAT (Robust Audio Tool), and VIC (Video Conferencing Tool), it offers data and screen-sharing tools including electronic notebook and TeleViewer. Fig. 4 shows a schematic representation of our MBone-based distributed collaboration.

In our tests, the audio/video information was interchanged among multiple endpoints simultaneously without using an MCU. The GeoVISTA Studio analysis was shared by different users via the TeleViewer, which provided a real-time view of a display area, a specific window, or the entire computer display to multiple participants (e.g., Keller & Myers, 1996). For example, in Fig. 4, users 2, 3, and 4 are synchronically sharing the computer screen of user 1, so that they can remotely view or control the data exploration while communicating about the exploration session via audio/video communication.

Our experiences thus far with MBone-based multicast group sessions have not demonstrated real advantages over Unicast collaboration. The learning curve to set up communications is rather steep and the user interfaces for open source toolkits are complex and difficult to understand. However, the network performance advantages of MBone cause it

---

**Fig. 4.** A diagram of the MBone-based distributed collaboration using GeoVISTA Studio analysis tools. User one’s screen is being shared by the other three collaborators. Multiway audio–video communication supports joint discussion in which User 1 helps the other users learn how to use the data analysis tools.
to be considered an important part of the infrastructure for collaboratory environments in the future (as technology becomes more usable and bandwidth increases).

For example, MBone is the strategy used to support audio/video streams in the Access Grid, which is considered to be the emerging standard for distributed, large, complex and persistent collaborations (Booth et al., 2002). Not only does the Access Grid provide a synchronous video/audio enabled interactive environment; it also facilitates a large-format different-place visualization environment supported by mass data storage and powerful data analysis and visualization technology (Relle, 2002; Tasker, 2001; Tierney, Johnston, Lee, & Thompson, 2000). We expect the Access Grid environment to improve HERO synchronous collaboration substantially, but only once it is available at all observatories in an accessible way – e.g., that does not require using a special facility across campus.

3.2. Asynchronous perspective comparison and consensus building with e-Delphi

As noted above, the primary goal of the HERO project is not to study human–environment interactions. It is to develop and assess data collection and analysis protocols for conducting such studies in a way that they can be compared from place to place and over time and in a way that makes integration of results into a comprehensive understanding of human–environment processes possible. One of the challenges facing distributed HERO collaborators is the need to make connections across team members’ specialized areas of expertise and to create expressions of group opinion. HERO collaborators approach human–environment interactions from a variety of perspectives, among them climate science, land use analysis, and public policy. Thus, interpretation of terms and concepts varies considerably and areas of agreement and disagreement about the drivers and effects of environmental change are rarely obvious. As a result, the process of deciding how to monitor environmental change and what parameters are critical in different locations can benefit from structured negotiation. Over time, team members argue their own positions, evaluate those of others, and gradually come to appreciate the breadth of a problem and the possibilities for its solution.

The Delphi method is one approach to structuring this exploration and distillation of expert opinion, and is a key approach used in our geocollaboratory. Developed at the Rand Corporation in the 1940s, the Delphi method was first applied to wartime forecasts by military planners (Dalkey, 1969; Linstone & Turoff, 1975). In principle the method works like an asynchronous focus group, run by a moderator who poses questions to a panel, synthesizes feedback, and guides the group toward its goal. This goal is not necessarily consensus, but the identification of patterns of belief, areas of agreement and divergence, or common opinions.

Traditionally, Delphi activities contain several rounds of participant input, each of which might prompt the participants to generate new ideas or refine, extend, vote on, or otherwise discuss responses from previous rounds. Typical Delphi activities are paper-based, conducted through the mail, although e-mail is also used (e.g., Bourque et al., 2002). Recent applications of the method in the environmental sciences include water resource issues (e.g., Nagels, Davies-Colley, & Smith, 2001; Taylor & Ryder, 2003) and climate forecasts (Tapio, 2002). Problems in these domains are generally complex and lack easily defined or readily agreed upon solutions. Collaborators in environmental science use Delphi to express and defend their beliefs about possible solutions with the aim of generating and refining a body of expert opinion. Environmental science and water resource problems can,
of course, benefit from real-time, same-place discussions. However, their complexity makes them well suited to an approach that allows participants to revisit an evolving discussion over weeks, months, or more, to do so from distributed locations, and to do so anonymously.

Anonymity is the first of four basic tenets of the Delphi method. The others are asynchronicity, controlled feedback, and statistical response (Turoff, 1971, 1972). Together, these four characteristics describe an approach to group communication that attempts to overcome some of the shortcomings of face-to-face meetings. Anonymity encourages participants to interact without the constraints of status relations or personality conflicts that can dominate in-person discussions. Generally, participants’ identities are known only to the moderator of a Delphi activity, and the moderator will often strip identifying information from responses to a question before forwarding them on to the other participants. Asynchronicity allows Delphi participants to engage in a discussion on their own schedules. A participant might choose to contribute only when the discussion approaches his or her areas of expertise, or after spending more time in reflection than would be possible in real-time meetings. Controlled feedback describes the moderator’s responsibility to shape the discussion. The moderator may choose how much information to present back to the panel, in what form, and how often. Through statistical response, participants’ contributions to a Delphi activity may be summarized quantitatively. They may be asked to vote on their support for particular ideas, with the results of this voting used to prioritize discussion in future rounds.

Paper- and e-mail-based Delphi activities have been successful, but are subject to a number of shortcomings. First, they place a significant managerial burden on the moderator. He or she must devote a substantial amount of time to gathering, organizing, editing, and returning participants’ responses. Second, the amount of time required to organize an activity can reduce the frequency with which the method is used or the number of problems to which it is applied. Third, participants do not have the benefit of organizing their view of the discussion in individually appropriate ways – they are restricted to the presentation format the moderator chooses, regardless of how well it suits their perspective.

We have developed a Web-based e-Delphi tool for the HERO geocollaboratory that acknowledges the Delphi method’s ability to organize group thinking while addressing the shortcomings outlined above (Fig. 5). The HERO e-Delphi system handles many organizational tasks for moderators, allows any number of simultaneous activities to be run, and increases the spontaneity with which participants can take part in the discussion. Through the e-Delphi web tool, any user can initiate a new activity – serving as the moderator – and can invite a panel of participants. The moderator sets a schedule for each round, provides instructions and guidance to the participants, and decides how much and in what form participants should be able to contribute. The moderator may also choose a level of anonymity to apply to an activity; participants may be entirely anonymous, or may be known to others by a user profile (containing information on expertise, affiliation, or experience) or by real name. The e-Delphi system uses the moderator’s settings to control access to the activity and to determine what information to mask or reveal. In contrast to traditional Delphi activities, participants can be allowed to see what others have contributed in real time, without waiting for it to be sent to and returned by the moderator. In addition, statistical response is updated continuously (in the form of vote results), allowing moderators or participants to keep abreast of the support for various contributions. By connecting through a standard Web browser, e-Delphi participants can contribute to an activity whenever and wherever they like.
An electronic environment for managing Delphi activities also supports HERO’s aim to understand the process of geographic knowledge construction and application. Managing all of a research group’s Delphi-style interaction through a single environment helps to build a rich knowledge base that reflects the evolution of a community’s thinking over time. Every contribution to an e-Delphi exercise is logged and a body of Delphi activities can be taken together to reveal trends in the concepts that are important to a group of researchers.
Pike and Gahegan (2003) have detailed how an evolving text corpus consisting of e-Delphi discussion can be mined (using other tools) for concepts at different levels of abstraction, and how the relationships among these concepts and the participants who discuss them can be visualized. This processing starts with simple term extraction, is followed by semantic abstraction (Mani, 2001) and calculation of concept similarity measures (Lin, 1995) to derive concept weights. These concept weights are then processed using Self-Organizing Maps (SOMs) to produce a series of 2D representations of the multidimensional concept space that evolves through a Delphi exercise (Fig. 6).

3.3. Long-term information sharing and knowledge development

The overall goal of the HERO project is to develop the infrastructure to support longitudinal study of the local and regional, human impacts of global environmental
change. To accomplish this, mechanisms are needed to archive and retrieve information and to capture and manage knowledge. The HERO team focused initially on tools for archiving information and has more recently developed and implemented knowledge capture and management methods and tools. Below, we outline the team’s approach to meeting each goal and briefly describe the methods and tools implemented.

3.3.1. Long-term information sharing: from e-notebooks to e-Workspace

Record-keeping and imagination-prompting are key activities for all types of scientific research. Traditionally, these activities are often supported by paper notebooks in which researchers compile sketches, project plans, raw data, and experiment procedures and results. Paper notebooks remain a primary means to compile information that supports scientific work because they are convenient, portable, and easy to use. However, paper notebooks are not good tools for supporting the work of large distributed science teams. Such teams need information encoding devices that enable information sharing as well as index and search capabilities (Myers, Mendoza, & Hoopes, 2001). An early effort to meet these goals was the electronic notebook (e-notebook) developed and implemented by researchers at the US Pacific Northwest National Laboratory, Environmental Molecular Science Laboratory (EMSL) (Edelson, Pea, & Gomez, 1996). Specifically, the EMSL team defined a scientific e-notebook as a web-based collaboration tool that supports sharing, viewing and analyzing documents, images, data, and other information among a group of collaborators (Edelson et al., 1996). Their web-based e-notebook, the Electronic Laboratory Notebook (ELN) is freely available to the scientific community and has been widely used (http://collaboratory.emsl.pnl.gov/intro/elnotebooks.html).

In the early stages of the HERO project, we adopted the EMSL-ELN for cross-project use. More recently, we have developed and implemented a shared workspace (called Codex) that provides more flexible tools to address HERO’s project-specific needs and those of the Geosciences Cyberinfrastructure Network (GEON) project (two members of the team are collaborators on GEON – http://www.geongrid.org/). Our experiences with each are outlined below.

3.3.1.1. EMSL electronic laboratory notebook (ELN). Developers of the EMSL-ELN set specific goals for their e-notebook that address the needs of laboratory science (with a focus on chemistry and biology). Key goals were that the e-notbook should be convenient to use, act as a permanent archive of all laboratory activities and their results, and implicitly encode the sequential order of those activities (Myers et al., 2001). The ELN implements a hierarchical chapter/page structure for organizing all information that is encoded by users. It includes search tools, automated e-mail notification of new entries, and sophisticated annotation capabilities. Time stamped notes (and subnotes) can be added using “entry editors” that support adding text, equations, whiteboard sketches, images, and arbitrary files linked to a location in an existing document. Notes can be viewed using external applications (e.g., a word processing package) or using applets or plugins (e.g., to depict a VRML or Flash note). The EMSL-ELN is now an open source product, distributed through sourceforge (http://eln.sourceforge.net/).

We experimented extensively with the EMSL-ELN in support of the HERO REU activities during year one of the project. The plan was for the e-notebook to serve as a central
resource through which project members could store and share distributed data and ideas. During this year, REU student research was focused on applying data collection and analysis protocols to a cross-site comparison of vulnerability assessments carried out at their four sites. To support this activity, students were encouraged to use the e-notebook to post concepts, methodologies, sketches, examples, and rationales as a record of ideas and to obtain feedback. We expected the e-notebook to be particularly useful for one task assigned to the REU students, the task of adapting the initial data collection and analysis protocol to make it more relevant to specific study areas. Our goal was a detailed record of how ideas and concepts changed over time. However, most of the REU students made only limited use of the e-notebook. In our debriefing with students at the end of the summer, they reported that the e-notebook (as configured then) had a counterintuitive interface design, its display of information seemed disorganized, and (because the e-notebook had no index) stored information was hard to locate. Students also pointed out the lack of any mechanism to enforce consistent naming and uploading of files; this coupled with the lack of index made it very difficult to compare information across research sites or observatories.

Two important lessons can be derived from this initial application of an e-notebook for support of human–environment science. First, the different conceptual models and styles of work inherent in different science domains demand different styles of information repository. The EMSL-ELN is designed to support typical laboratory science where the laboratory notebook is an inviolate permanent record of those activities and results. To fit this domain, the EMSL-ELN only supports additions and annotations, not changes. In contrast to laboratory science, which follows already develop protocols that require careful and systematic accounting, the HERO project is focused on developing these protocols, thus needs a more flexible information repository that does not discourage sharing of ill-formed ideas for input by others. Second, building a repository of shared knowledge takes effort, particularly when the technology to support it is new and fragile. It became clear that the benefits of tools that support data and knowledge resources sharing need to be carefully articulated and demonstrated or they are likely to be bypassed as more trouble than they are worth.

3.3.1.2. Codex. Our experiences with the EMSL-ELN during year one of the project were instrumental in the conceptualization, design, and implementation of the Codex web portal. In place of a notebook metaphor, Codex adopts a workspace metaphor that links together a nexus of science activities, including people, methods, data, and concepts, into a searchable web of relationships. The HERO Codex has been designed to provide users with five entry points into a collaborative scientific workspace, through concepts, tasks, places, groups, tools and files (data resources) (Fig. 7). The conceptual approach that underlies Codex is introduced in a pair of earlier papers (Pike & Gahegan, 2004; Pike, Yarnal, MacEachren, Gahegan, & Yu, 2005). Here we sketch an outline of Codex capabilities and focus on experiences in using it to support REU activities within the HERO project.

Design and implementation of Codex has been driven by the scientific infrastructure research goals of the HERO and the GEON projects. Specifically, those goals emphasize the need to develop methods that capture both scientific knowledge and the process by which knowledge is created. To address the latter, the HERO Codex attempts to
contextualize the use of HERO resources by logging the situations that surround their use (including aspects such as who used them, where and when, with what other resources, and so forth). Codex supports groups to provide a mechanism for both long-term and ad hoc teams to share content, tools, etc. Groups can be generated as needed. The most important uses of Codex have been as an information repository (for data, reports, in-progress papers) and as the portal to the web version of ConceptVISTA tools that have been used for collaboration on building detailed concept maps and comparing concept maps for the different HERO sites. A demo of the Codex ConceptVISTA tools is accessible from: http://flatbox.geog.psu.edu/codex/jsp/help/about.jsp.

contextualize the use of HERO resources by logging the situations that surround their use (including aspects such as who used them, where and when, with what other resources, and so forth). Codex supports groups to provide a mechanism for both long-term and ad hoc teams to share content, tools, etc. Groups can be generated as needed. The most important uses of Codex have been as an information repository (for data, reports, in-progress papers) and as the portal to the web version of ConceptVISTA tools that have been used for collaboration on building detailed concept maps and comparing concept maps for the different HERO sites. A demo of the Codex ConceptVISTA tools is accessible from: http://flatbox.geog.psu.edu/codex/jsp/help/about.jsp.

Codex also contextualizes concepts, whether they are defined by a user or repurposed from concepts and relations created by others. For example, an investigator studying flood risk in one region can find out how peers (in other places and at earlier times) defined similar concepts. The investigator might find a description of flood risk that includes data local to the place where the description was created (such as USGS stream gauge files). Assuming that these data are not suited to his/her needs, he/she modifies the description to incorporate new data, and may add or delete some of the description’s concepts to fit the circumstances of the location being studied better. Then, by contributing this modified knowledge structure back to the community through Codex, two benefits result: first, it becomes possible to trace the evolution and adoption of geographic knowledge over time and across space; second, overlap between multiple conceptual structures can be detected and used to indicate agreements or disagreements among places and across geographic scales.
Overall, representing information in a semi-formal manner gives researchers the flexibility to describe a wide range of relationships, including those that are vague or inconsistent. It also removes the need for all proposed concepts to be based on some formally agreed base language, some form of foundational ontology that does not yet exist for human–environment interaction. Similarly, maintaining continuous records of a concept’s modification and use allows users to situate concepts historically and allows users to explore their development. To support understanding of the context of scientific concepts, Codex includes tools for visual display and exploration of the provenance of any resource (either abstract concept or tangible data file).

As a complement to exploring a concept itself, Codex provides mechanisms for exploring a concept space as a social network – revealing relationships between researchers who used a resource in their own work. Codex assumes that different users may rely on different data, methodologies, or theories and have different expertise and experiences. To enable joint work with this diversity as a starting point, Codex provides tools to search for similarities across contexts, including researchers, locations, times, goals, methods, and so forth. This is supported by mechanisms that translate visual depictions of knowledge – typically created by the scientists themselves – into computational representations in OWL.

By enabling users to link concepts, data, and methods into in silico experiments, Codex supports a range of information tasks; these include:

- recording the construction and evolution of knowledge within a defined community,
- finding and accessing existing knowledge components that suit a current task,
- discovering and visualizing concept relationships,
- finding data sets and tools that have been used to solve similar problems.

3.3.2. Capturing and sharing knowledge: ConceptVISTA

The knowledge infrastructure goals of HERO are supported in Codex through the formal representation of concepts along with tools that support capture, sharing, evolution, and comparison of concepts through concept maps. To support these activities, Codex uses an instantiation of ConceptVISTA, a concept browser/ontology editor, built initially as a stand alone application within GeoVISTA Studio (Gahegan, Dai, Macgill, & Oswal, 2003). In contrast to Codex, ConceptVISTA is designed more as an educational aid or quick sketch-pad of ideas that requires no network link back to a server. We have used ConceptVISTA in group meetings within both the HERO and GEON projects to capture concept maps during the course of a discussion. These help to articulate the complexity (or simplicity) of protocols visually as they are proposed and to form part of the permanent record of the meeting. Again, the representation of knowledge is kept deliberately informal, so that expert scientists are not restricted by the need for formal consistency and closure as they brainstorm ideas. We are currently investigating ways to migrate this informal knowledge, without loss of content or context, to the formal, computable structure afforded by ontologies and their related inferential systems.

Two examples of ConceptVISTA applications within HERO are given below. The first is a snapshot captured during an all-hands project meeting in Arizona, 2004, early in the development of a HERO protocol for assessing vulnerability to climate change (Fig. 8). Note that the notion of vulnerability has three main constituent parts: exposure, sensitivity and adaptive capacity. Each component is formed from progressively less abstract
components (nodes furthest from the central vulnerability node). Note that the concept map supports integration of geographic concepts (physiography, areal extent of stressors, distribution) as well as temporal concepts (frequency, age, duration). Thus, concept maps support encoding of geospatial and space–time knowledge. ConceptVISTA supports connections between knowledge and data through direct links from nodes in the concept map to supporting resources (e.g., databases, tables, maps, images).

The second example depicts a screenshot from a concept map showing the 10 most significant events that have changed the Pennsylvanian landscape, enhanced with domain content shown in the right-hand panel (Fig. 9). The 10 most significant events form a core of concepts (diamond-shaped) grouped around the Central Pennsylvania (PA) concept (large square). The concept currently selected (or ‘in focus’) is agriculture (shown in white), and it provides the description and images shown in the right-hand panel. The content shown here was manually captured in large part from HERO working papers and provides a rather interesting alternative interface to the material when compared to an academic text that must be read linearly, from start to finish. By contrast, this visual interface provides navigation capabilities similar to a hypertext document, except with an adaptive display that reacts to the user’s need for more general or more specific information. By using the ‘locality’ slider control (bottom right in either figure) the concept structure can be presented progressively – starting with the concept currently in focus and expanding through the different degrees of connections until all concepts are shown.
4. Discussion and future work

As noted in the background section, past collaboratory efforts have focused on supporting distributed laboratory science and remote use of scientific instruments (e.g., telescopes, magnetic fusion devices, etc.). Our research, reported here and elsewhere (see MacEachren, Gahegan, & Pike, 2004; Pike, Ahlqvist, Gahegan, & Oswal, 2003; Pike et al., 2005), has focused on extending the collaboratory concept and technologies to enable distributed social science focused on understanding human–environment interaction. Key challenges in that effort have included:

- support for thinking and knowledge construction related to ill-defined tasks associated with developing new protocols for data collection that includes both qualitative as well as quantitative data,
- support for both experienced research scientists and students learning about the process of carrying out social-environmental science research,
- support for geographic representation in maps, image, and other forms (e.g., place descriptions, concept maps),
- support for scientists who do not have regular access to well-equipped research laboratories with technical staff.
Our experience deploying these collaboratory tools for use in a real-world multisite research project has demonstrated that each fulfills a clear need and has unique benefits to the research community. The tools support collaboration at a range of time scales (from short-term conferences to long-term archiving) and communication across a spectrum of interaction types (from augmented face-to-face meetings in video conferences, through anonymous discussion and synthesis in Delphi activities, to shared work records in an online portal). Having a suite of tools – rather than a one-size-fits-all solution – helps a team tailor their choice of tool to meet the requirements of a particular situation, rather than adjust their work style to fit a tool’s capabilities.

Development of useful and usable geocollaboratory methods to enable different-place science work has required an iterative process of implementing, assessing, and refining both methods and strategies for integrating them into work. Our attempt to use geocollaboratory tools early in the project had mixed results. As detailed above, the initial e-notebook used in the project was found to be a poor match to project needs due to its basis on the rigid, physical science laboratory notebook metaphor. All of the various video conferencing methods used have been considered useful by team members, but each has had different technical implementation issues to address. In general, video conferencing linked with desktop sharing has proven to be more useful for small, informal project meetings between individuals (or very small groups) at pairs of sites than for large project meetings. The e-Delphi tools have been effective for building shared understanding among senior scientists on the project but have been less successful for similar sharing among REU students. The latter have indicated some frustration with the structured process of Delphi exercises and with the time they require to complete; it is also likely that the topics we have asked REU students to explore with e-Delphi were not well matched to their level of background knowledge and domain expertise.

Many of the collaboratory methods and tools we have implemented and applied would be equally useful for support of non-geographic science. The HERO geocollaboratory is geographic not so much because it emphasizes traditional geographic representation (e.g., maps and images) but because of the fundamentally geographic problems it is applied to – ones that focus on understanding and comparing places. The HERO geocollaboratory does explicitly support the integration of geographically referenced data and knowledge, but typically in forms other than GIS layers. Using Codex, data/knowledge resources that are contributed can be integrally tagged with the location(s) to which they apply and/or the location at which their creator works. As a result, we can achieve some measure of understanding about the nature of overlap and idiosyncrasy in how concepts are defined in different places. “Where” is an inherent part of the resource and collaborator metadata in Codex. This distinguishes our approach from other approaches to scientific collaboratories. By default, everything in Codex becomes “geographic” information – it is not necessarily spatial data (in the GIS sense), but it is spatially referenced. The novelty of our geocollaboratory is that we geographically reference more than “features” – concepts, people, data resources, and all other encoded entities are “spatially referenced” in Codex too.

Our current geocollaboratory implementation has emphasized low-cost, widespread technologies to increase ease of adoption by collaborators. While these tools will likely remain the backbone of collaboratory efforts in the human–environment sciences because they require minimal technological infrastructure, some types of collaboration require high-performance computing over high-speed networks. An Access Grid enabled meeting room with large, high resolution display now exists at Penn State and is well suited to
supporting remote, large group meetings that share maps and images. In the near future, efforts of the HERO synchronous geocollaboratory will focus on testing and using the Access Grid to improve our existing geocollaboratory tools.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants No. BCS-9978052 and ITR (EAR)-0225673 (GEON). We wish to acknowledge the more than three dozen REU students who have participated in the HERO project and whose use of the methods and tools described has helped to assess and refine them. In addition, we thank Biliang Zhou for his help in conducting some tests of our synchronous collaboration tools, Jason Heffner from Penn State’s Information Technology Services for help with video meeting technology, and James Myers from the US Pacific Northwest National Laboratory, Environmental Molecular Science Laboratory for his help in adapting the EMSL-ELN to HERO activities. Finally, we acknowledge support from the IBM Academic Initiative (formerly called the IBM Scholars Program) that supports use of IBM products (at no charge) for teaching and research – this program provided use of Lotus SameTime and technical support in setting up our SameTime server.

References


