Collaborative Design of an Oceanographic Event Logger

Brian Lindseth
Department of Sociology
Science Studies Program
University of California, San Diego
blindseth@ucsd.edu

Karen Baker
Graduate School of Library and Information Science
University of Illinois, Urbana-Champaign
karensbaker@gmail.com

ABSTRACT
We report on an in-progress project to design a new field instrument used to link data collection and preservation practices across the gaps separating groups of oceanographic scientists engaged in interdisciplinary collaborations. Oceanographic cruises provide an important means for collecting data as an input for specific scientific investigations and for longer term efforts to make a wide range of environmental field data available for future scientific investigators. For both of these tasks, there is the need to perform the invisible work of identifying and resolving potential gaps and consistency issues in the data in order to make the data usable. The event logger system was designed to address these issues by associating every measurement on two organization's scientific cruises with a latitude, longitude, and time stamp. It is the finding of this ethnographic analysis that the event logger system has been successful to the extent that the design process has been able to incorporate a diverse range of voices into an open and collaborative activity.

Author Keywords
Collaborative design, information infrastructure, invisible work, oceanographic research.

ACM Classification Keywords
H.5.2 User Interfaces: User-Centered Design, Prototyping; H.5.3 Group and Organization Interfaces: Computer-supported Cooperative work;

General Terms
Design, standardization.

INTRODUCTION
We provide an account of an event logger system deployed recently on a series of oceanographic research cruises. Each of these cruises brings together a variety of scientists interested in collecting data in the waters off the coast of Southern California to investigate a wide range of research questions. In addition to the data being used to answer specific research questions, there is an imperative to render the data interoperable across the contexts in which it was collected so that it can provide a resource for future, potentially larger scale research questions.

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.

At the level of specific investigations, measurements must be associated with a specific time and place in order to be meaningful. Seemingly small differences in establishing the exact sequence and location of measurements become progressively more difficult to reconcile as the ship data return to land and are carried by investigators into their labs. Bringing attention to this invisible work and enlisting a broader range of actors to the task of ensuring the accuracy and usability of data at the level of specific scientific projects has been a design target for the event logger.

Another design goal of the event logger system has targeted issues inhibiting efforts to bring together data collected by different scientific organizations as resources for future scientific collaborations. As different organizations' cruises employ very different kinds of sampling strategies, measurements made in different contexts are tagged in different ways—sometimes associated with a fixed grid of stations and sometimes with moving masses of water that are followed by the ship. In the contexts of these different kinds of cruises, the event logger provides a common way of locating measurements in time and space in a shared, annotated log so that data generated by different organizations can be rendered commensurable more easily. The event logger was designed in order to minimize under-recognized post cruise data management work and to facilitate efforts to render data collected on these cruises interoperable across organizational divides.

DESIGN: SETTING, SYSTEM AND PROCESS
It is very important for measurements generated within a given team of scientists to be linked with a definite time and place where they were taken in order for these measurements to be usable. Traditionally, this information has been recorded by hand by different cruise personnel according to different conventions. Further, details of time and place can sometimes be omitted in the chaotic conditions of taking measurements in the middle of storms at sea or amidst make shift efforts to address equipment malfunctions that threaten to delay sampling schedules and prevent the collection of valuable data. Even under the best of conditions, a cruise member documenting the time of a

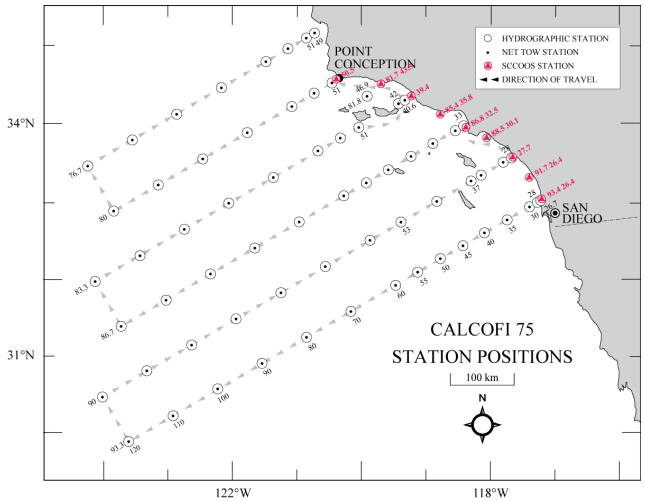
measurement could draw the time from one of two global positioning satellite systems (GPS)—potentially configured differently in the GPS' receiving equipment—or from any one of a bevy of unsynchronized shipboard computer clocks. Even with all notations in place then, discrepancies in the timing of measurements can emerge because of the availability of multiple possible inputs.

Traditionally, the work of addressing these gaps and discrepancies takes place after the cruise ends among graduate students, data managers and cruise and scientific support personnel. For many of the scientists leading the research, this work is often invisible [2,6,7,18,19]. As cruise experiences fade in memory and resurface less frequently in conversational encounters with time, such discrepancies become progressively more difficult to resolve as the links between measurement data and the contexts in which they were taken become more and more attenuated. One data manager commented on many scientists' tendency to take the work of resolving gaps and inconsistencies in the data for granted: **"**All that. . . metadata work [of] going to many people [who had been on the cruise] to resolve inconsistencies, that step's often taken for granted, and it gets harder as collaborative science grows. And yet, it's not budgeted in terms of the pressure from the PI. 'Well certainly you can get this processing done in a week.' Whereas they're not counting the day and a half that it takes to do that bookkeeping that hasn't been done." In the same interview, this data manager contextualized this problem in terms of the growth of scientific collaborations: "In the past they [specific people working in data management roles] had it done because they used to work with investigators for years. And they had all these coordination mechanisms passively agreed upon. When you have a new cruise like this with a new team, none of that is in place. And so the burden on someone like L. handed a notebook [full of cruise data and asked to resolve problems in the data] like that [when she] wasn't on the cruise, is tremendous." The event logger was designed in order to facilitate this invisible, post-cruise work by associating each measurement with a definite time and place as the measurement is made on the cruise.

The scope and impact of this kind of post cruise work in making the data usable grow when more than one team of scientists is interested in using the same data. Oceanographic cruises are composed of investigators with different disciplinary and institutional backgrounds and with a variety of research interests brought together by a shared interest in describing and understanding the marine system off the California coast. Perhaps unsurprisingly, the data collected by different scientific teams and particularly across organizational boundaries require extensive work to be rendered commensurable. Research focused on the social dimensions of large scale Cyberinfrastructure projects in the U. S. context and eScience projects in the European context [2,3,4,9,17] describes precisely this kind of work.

Assisting in rendering data generated on cruises interoperable across divisions separating larger scale organizational contexts is second design goal for the event logger. Differences in the sampling strategies of California Cooperative Oceanic Fisheries Investigations (CalCOFI) cruises and a subset of the Long Term Ecological Research Network's (LTER) California Current Ecosystem (CCE) cruises provides an illustration of the kinds of problems encountered in rendering data interoperable across organizational boundaries.

CalCOFI cruises build on a sampling pattern that was first established in the years following World War II and that was based on the original mission of the CalCOFI project--to understand the collapse of sardine populations. There are a number of stations extending out from the California coast in a grid, the majority of stations roughly 40 nautical miles away from the next station. Typical cruises in the summer and fall collect samples at, or "occupy," seventy five of these stations over eighteen days. For CalCOFI participants, stations are generally referred to by the number of the line running perpendicularly away from the coast and then the number of the station along that line. 90.60 is station 60 along line 90. Sometimes, however, these stations take on less formal names as part of the experience of each cruise. In order for the data collected on CalCOFI cruises to be meaningful in comparison to other samples that are part of a given study, it is very important that they be associated with the precise time and location—or formal station name—where they were taken. Established much more recently, the Long Term Ecological Research Network's (LTER) California Current Ecosystem (CCE) project began joining CalCOFI's cruises in 2004 as part of an effort to build on the decades of data collected by CalCOFI. In addition to accompanying CalCOFI scientists on cruises following a station based sampling strategy, the CCE site engages in "process" cruises that do not follow the long standing station-based sampling pattern established by CalCOFI. Instead these CCE cruises follow masses of water



CalCOFI station based sampling pattern.

for several days and take a variety of samples in a series of four to six sampling periods or "cycles."

By associating each measurement—whether from CCE process cruises or CalCOFI station based cruises—with a precise latitude, longitude, and time, the event logger was designed to render these measurements interoperable across these organizations and sampling strategies. In the process, the event logger could provide a mechanism for storing measurements that might have fallen through the cracks of CalCOFI's station based approach to sampling. One CalCOFI cruise participant noted the inaccessibility of data that did not fit neatly into the organization's station-based sampling with its standard station names: "CalCOFI's always been focused on stations, so anything between the stations has been ancillary or extra. And it's not—it's not reported by CalCOFI or stored in their databases for public consumption. You know, it's all on a DVD somewhere. . . but it's not easy to get at".

What does the event logger system look like? The system is made up of two or more tablet computers—at least one placed in the lab area of the ship and one on the bridge. Both scientific support personnel and bridge personnel interface with a custom program on the computers in order to associate planned measurements with latitude and longitude and a time stamp as the measurements take place. An event log is produced as an output, a list of each measurement that has taken place on board the cruise. It is a sequential, digital record in a standard format that can be accessed by all cruise participants. Together, the event logger and the event log were designed to span the flow of data - from measurements in the field to specific scientific investigations and a shared data repository. The standard time and location allow for translation of measurements across organizational boundaries, and a unique identifier (the event number) allows commensurability to be enacted at the level of the database, establishing relations across research organizations and ultimately enabling queriability. In this context, the event logger was designed as a way of distributing data management work [11,12]. Instead of relegating the work of cleaning up the data to post cruise data users, the work of "keeping the data clean" – organizing and relating the data - begins at the source of the data itself in measurements on board cruises. In extending post-cruise reconciliation efforts into the cruise, the event logger functions to distribute this work over time and a wider set of actors [10,11,12,14].

Sampling pattern for a LTER CCE 'process' cruise superimposed on a map of CalCOFI stations.

FINDINGS

The central finding of this project so far has been the importance of an open and collaborative design approach to the ongoing success of the event logger as a socio-technical system [17]. Although the event logger was designed as part of an effort to highlight and distribute invisible data management work, it was initially defined primarily in technical terms as a 'lightweight' system impacting a relatively narrow range of cruise participants. As the system was implemented and met with occasional resistance and breakdowns, the design team quickly adjusted its approach and attended in greater detail to the working environments of the cruise [8,10,11,12,13], soliciting the feedback from a much wider range of cruise-based personnel than were initially considered. In incorporating a large number of cruise personnel into the study as designers and not merely as less involved support personnel [6,7,14], the event logger system gained support and achieved a greater degree of success. In order to take advantage of this larger circle of designers, a greater degree of the system's functionality was made configurable. On one cruise this allowed greater flexibility in associating measurements with organizational settings.

The initial resistance of some bridge personnel was unanticipated and provided a prompt to solicit interviews with ship-based personnel and initiate an ethnographic study in order to determine how the system was being used, and not being used, in practice. One of the things we found was that the event logger was initially disproportionately supported by a few scientific support personnel already convinced of the system's value. One research participant had the following to say about the crew's initial reaction to the event logger: "So first it was rough going, and the crew. . .didn't seem very happy with it at first. It took them a while to adjust to it because it's not replacing something that they've already been doing. It's just another added feature. . . Ultimately (chuckle) a few of the mates are like so are you taking the event logger with you? (chuckle) So it's like, oh sorry guys. I mean it is a little bit more of a extra duty for them." Although this input was not the glowing review that the design team had hoped for, it provided a key opportunity to include the voices of a range of cruise personnel into the design of the event logger as a more open and collaborative process [1,6,7,14].

One cruise member, for example, suggested that the event logger become more integrated with the ship's already existing logging mechanism: "So if it can ultimately get integrated with a bridge log, that would be easier to do or instead of having for them to write stuff out, maybe if there was a keypad up there and then like they can add in their

own comments, like oh we're having a fire drill today or something like that." Although the functional division separating the role of the bridge from the rest of the ship would make implementation of this suggestion difficult, it called attention to the need to approach bridge personnel as an active part of the event logger system—effectively widening the initial design conception of the people affected by the event logger system. Once the crew were approached directly as part of a dialog about the future place of the event logger, they have been much more responsive to the positive role that it can play for scientists, scientific support personnel, and data managers.

In implementing the insights gained from a more open and collaborative approach, the event logger has become a success story. Crew member responses to occasional technical issues provides an important indicator of the impact of adopting a more open and collaborative approach to the design of the event logger. Whereas early technical malfunctions were sometimes seen as an opportunity to ignore the event logger for the remainder of the cruise, more recent problems have been met by efforts to improvise fixes. When problems were encountered with the tablet computers on one cruise, the event logger software was installed on a laptop that became the de facto interface for the event logger system. Now the event logger is not only an accepted component of these cruises,

In going forward, the design team hopes to resolve the impact crew turnover has on the event logger system. Turnover is not only one of the biggest remaining issues for the event logger system but also a reminder of the importance of viewing the event logger as a system that is both technical and social in nature. As new crew members focus initially on their duties, overlap in staff has been important in introducing new personnel to the event logger system in ad hoc training sessions initiated by other cruise participants. As the design team formulates back ups to these ad hoc training sessions, it will be important to include the new crew members as potential participants in an open and collaborative design process.

ACKNOWLEDGEMENTS

Thanks to Geof Bowker, Florence Millerand, David Ribes, and members of the LTER and Ocean Informatics communities at the Scripps Institution of Oceanography.

REFERENCES

1. Baker, K.S., Jackson, S.J., and Wanetick, J.R. Strategies supporting heterogeneous data and interdisciplinary collaboration. Proceedings of HICCS 06 (2005).
2. Baker, K.S., and Millerand, F. Articulation work supporting information infrastructure design. Proceedings of HICCS 07 (2007).
3. Bietz, M.J., Baumer, E.P.S., and Lee, C.P. Synergizing in Cyberinfrastructure Development. Computer Supported Cooperative Work 19 (2010), 245-281.
4. Bowker, G.C., Baker, K.S., Millerand, F., and Ribes, D., Toward Information Infrastructure Studies. International Handbook of Internet Research. J. Hunsinger, L. Klastrup, and M. Allen, Eds. Singer, New York, (2010).
5. Bowker, G.C. Memory Practices. MIT Press, Cambridge, MA, (2006).
6. Bratteteig, T. Making change: dealing with relations between design and use. PhD Thesis, University of Oslo, Norway, (2003).
7. Clement, A. Looking for the designers: Transforming the 'invisible' infrastructure of computerised office work. AI & Society 7, 4 (1993), 323-344.
8. Dourish, P. Where the Action Is. Cambridge, MA: MIT Press, (2001).
9. Edwards, P., G. C. Bowker, S. J. Jackson, and R. Williams. An Agenda for Infrastructure Studies. Journal for the Association of Information Systems 10, 5 (2009), 364-374.
10. Halverson, C. A. and Ackerman, M.S. 'Yeah, the Rush ain't here yet – Take a Break': Creation and use of an artifact as organizational memory. Proceedings of HICCS 03 (2002).
11. Hollan J., E. Hutchins, D. Kirsh. Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. ACM Transactions on Human Computer Interaction 7, 2, June (2000), 174-196.
12. Hutchins, E. Cognition in the Wild. MIT Press, (1995).
13. Kaptelinin, Victor, Nardi, B. Acting with Technology: Activity Theory and Interaction Design. Cambridge, MA: MIT Press, (2006).
14. Kanstrup A.M. and Bertelson, P. Participatory IT Support. PDC 2006 – Proceedings of the 9th Participatory Design Conference (2006), 87–94.
15. Lee, C. Boundary Negotiating Artifacts. Computer Supported Cooperative Work 16 (2007), 307-339.
16. Ribes, D. and Lee, C. 2010. Sociotechnical Studies of Cyberinfrastructure and e-Research. Computer Supported Cooperative Work, 19, (2010), 231-244.
17. Star, S.L. and Ruhleder. K. Steps towards an ecology of infrastructure. Computer Supported Cooperative Work, Chapel Hill, (1994), 111-134.
18. Star, S.L. and Strauss, A. Layers of silence, arenas of voice: the ecology of visible and invisible work. Computer Supported Cooperative Work 8 (1999), 9-30.
19. Suchman L. Making Work Visible. Communications of the ACM. 38, 9, September (1995), 56-64.