

Project Summary

This project uses ethnographic methods to provide a situated social and organizational comparison of three scientific cyberinfrastructure projects deploying different approaches to achieving data interoperability. The three projects are GEON (<http://www.geongrid.org>), which is using a national distributed storage broker to create data sharing across multiple disciplines through developing shared ontologies; LTER (<http://lternet.edu/>), which is using metadata standards to federate data across single disciplines; and the Ocean Informatics team at Scripps Institute of Oceanography which is using metadata to share data across multiple projects within a discipline.

1. Intellectual Merit of the Proposed Activity

As the new scientific cyberinfrastructure is emerging, a central question being posed is how to share data across multiple distributed organization and social contexts. While there have been a wealth of suggestions for technical fixes for this pressing concern (particularly important since some of the great political questions of our day, such as preserving biodiversity and developing a sustainable relationship with our environment pivot on the ability to federate data across organizational and disciplinary contexts), there has been little study – and no comparative study – of the organizational and social dimensions of differing interoperability strategies. Our working hypothesis, drawing on research in the field of social informatics over the past fifteen years, is that the creation of a common shared data infrastructure entails complex negotiations relating to the relative institutional weight of the different actors (institutions have a range of motives for subscribing or not to interoperability strategies), the nature of their disciplinary organization (in particular reward structures; openness to interdisciplinary work; history of use of large datasets) and the nature of their domain work (degree of commitment to long-term data storage and re-use; decay rate of data over time; need to draw on large federate datasets). Through this study, we will develop a grounded understanding of the organizational complexity producing shared scientific cyberinfrastructure.

2. Broader Impact Resulting from the Proposed Activity

The development of scientific cyberinfrastructure is vital for this country's future economic prosperity and for its ability to respond to key policy issues with scientific and technical dimensions. The project will facilitate understanding at the level of science policy of organizational and social dimensions of the development of shared cyberinfrastructure. We will produce a policy white paper on data communities and scientific cyberinfrastructure. It will suggest guidelines for the ongoing formative evaluation of infrastructuring activities. Further, with the continuing development of scientific cyberinfrastructure, there is a need to develop educational programs – both inreach and outreach – which sensitize domain scientists, computer scientists and science policy workers to social and organizational issues. Bowker will develop a graduate course in the development of cyberinfrastructure, as a centerpiece to his push for a Masters level program in cyberinfrastructure. We will develop a website modeled on Chip Bruce's Inquiry page (<http://inquiry.uiuc.edu/>) which will enable us to share or results first within the communities we are studying and then across communities. This will provide the kernel for a resource to be developed as a site for researchers and practitioners in the emergent field of scientific cyberinfrastructure to share findings and best practices and to engage in collective problem-solving.

Introduction

This project uses ethnographic methods to provide a situated social and organizational comparison of three scientific cyberinfrastructure (CI) projects deploying different approaches to achieving data interoperability. The three projects summarized in Table 1 are the Geosciences Cyberinfrastructure Network (GEON; <http://www.geongrid.org>), which is using a national distributed storage broker to create data sharing across multiple disciplines through developing shared ontologies; the Long-Term Ecological Research Network (LTER; <http://lternet.edu/>), which is using metadata standards to federate data across single disciplines; and the Ocean Informatics team at Scripps Institution of Oceanography, which is taking a community-driven approach and designing a local metadata standard to bridge key data collections to a national standard.

We are moving from an era of single purpose information systems, designed for specific agencies or disciplines into an era when the most significant political and scientific questions of our time can only be dealt with through the creation of a single cyberinfrastructure with multiple constituencies. We cannot make good decisions about fresh water supplies without drawing upon the skills of hydrologists, geologists and conservation biologists (multiple scientific domains) on the one hand, and the distributed databases of assorted governmental agencies (multiple organizational contexts) on the other. The problem of designing data repositories and modeling tools for multiple constituencies is radically different from the traditional computer science application targeted at a single well-defined community-of-practice (Lave, 1988). In the latter case, the range of uses for a data source are indexed by the needs of a user community operating within a broadly shared set of paradigms (analytic tools, measuring instruments and theoretical questions). In the case of design for cyberinfrastructure, this is typically not the case: data collected for one reason (climate modeling perhaps using satellite photos) might well be deployed for a completely different one (as a marker for vegetation cover and hence surrogate for fauna). Data interoperability is a *sine qua non* for the development of effective cyberinfrastructure. It is not enough to offer access to publicly funded data (Arzberger et al, 2004)—the data has to be usable as transparently as possible in a range of different informational settings.

Table 1. Communities and Community Elements

Community Element	GEON	LTER	OCEAN INFORMATICS
Project Inception	2002	1980	2004
Domain Focus	geosciences	ecology	oceanography
Partner Organizations	USGS, DLESE, SDSC	NCEAS, SDSC	IOD, SIO, JGOFS, SDSC
Initial Vision	temporal&spatial scales	long-term data	user-centered infrastructure
Data Organizational Unit	national distributed storage broker	federated sites	local storage
Standards Type	ontology	metadata	metadata
Scope of Standards	interdiscipline	discipline	cross-project
Standards Implementation Unit	community	site	division
# Lead Scientists	15	24	4
# Working Community Participants	1500	1300	300
Technological Unit	San Diego Supercomputer Center	sites with Network Office	division computer center
Organizational Unit of Work	2 regional test beds (rocky mt, atlantic)	24 biomes	3 data collections
Education	DLESE	Schoolyard LTER	local inreach & outreach
Social Science Role	observer	participant observer	observer participant
Info Mgt Driver	storage & access system	data sharing practice	community building
Info Mgt Goal	access national	site science	technological awareness
Info Mgt Environment	open-source	open-source	open-source
	in-house developed	partner developed	off-the-shelf
Info Mgt Organizational Unit	grid	Network Information System	web portal

The National Science Foundation’s vision for cyberinfrastructure is that it constitutes a fully integrated interdisciplinary tool providing access to domain scientists, educational institutions and to the lay public (NSF-AC-ERE, 2003; Futrell, 2003; NSF-AP, 2003). We propose through a comparative study to build a framework for understanding organizational and social dimensions of designing for multiple constituencies. We build on a body of literature that has emerged within the fields of science studies and social informatics to design a research plan which will enable us to:

- Systematically analyze the process of the negotiation of interoperability across multiple constituencies through an exploration of differing strategies from three distinct communities;
- Inform the design process through regular feedback with the design community modeled on the principles of participatory design;
- Develop a curricular program and a web-based ‘Cyberinfrastructure Page’ for the CI community.

Background

As the new scientific cyberinfrastructure is emerging, a central question being posed is how to share data across multiple distributed organization and social contexts. While there have been a wealth of suggestions for technical fixes for this pressing concern (particularly important since some of the great political questions of our day, such as preserving biodiversity and developing a sustainable relationship with our environment pivot on the ability to federate data across organizational and disciplinary contexts), there has been little study – and no comparative study – of the organizational and social dimensions of differing interoperability strategies. Our working hypothesis, drawing on research in the field of social informatics over the past twenty years (Kling and Scacchi, 1982; Bowker et al, 1997; Kling, 1999; Brown and Duguid, 2000), is that the creation of a common shared data infrastructure entails complex negotiations relating to the relative institutional weight of the different actors, the nature of their disciplinary organization, and the nature of their domain work. Institutions have a range of motives for subscribing or not to interoperability strategies, individuals within an organization are subject to particular career and reward structures, organizational culture speaks to the openness to interdisciplinary work or the history of use of large datasets; particular domains have specific degrees of commitment to long-term data storage and re-use; decay rate of data over time or need to draw on large federated datasets.

For this study, we adopt Bransford’s notion of environments (NRC, 2001) and portray cyberinfrastructure in Figure 1 at the union of spheres, in this case mediating for domain science, computer science and technology, and education. CI mediation is informed by communication and science studies as well as information management (IM) and information science. It provides a frequently invisible support for community in general. Through this study, we will develop a grounded understanding of the organizational complexity producing shared scientific cyberinfrastructure. Further, we will develop a three-pronged educational project that creates an interdependent set of broader impacts by– providing ways of *reading* cyberinfrastructure at the secondary school level (a key tool for the informed citizen of the twenty-first century); ways of *creating* it by developing a curricular kernel for merging science studies and

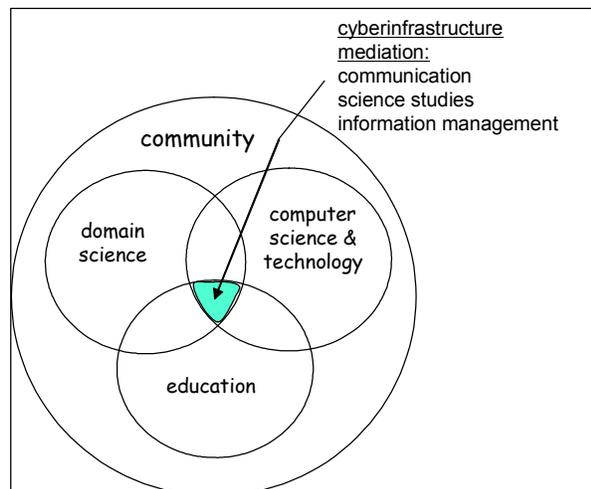


Figure 1. Conceptual schematic representing the work of cyberinfrastructure in mediating within the context of community among the domain sciences (ecology, geosciences, and oceanography), computer science & technology, and education.

information science to train the next generation of cyberinfrastructure information managers; and finally ways of *interacting* with it by developing a ‘Cyberinfrastructure Page’ for the CI design community.

We will work from an expanded version of Star and Ruhleder’s (1996) definition of the salient features of infrastructure in order to bound and clarify the term: (i) *embeddedness*, (ii) *transparency*, (iii) *reach or scope*, (iv) *learned as part of membership*, (v) *links with conventions of practice*, (vi) *embodiment of standards.*, (vii) *built on an installed base*, (viii) *becomes visible upon breakdown*. Something that was once an object of development and design becomes sunk into infrastructure over time, it becomes more solid, less malleable.

Just as cyberinfrastructure itself is a novel sociotechnical form, so too is the study of cyberinfrastructure building. This said, cyberinfrastructure carries many historical lineages to previous forms, most specifically to information infrastructures, and infrastructure more generally. These phenomena within the fields of science and technology studies have received considerable study within science and technology studies, communication and organizational theory. By drawing together these resources it will be possible to form an integrative understanding of the development of what it takes to make data interoperable; and what developers, domain scientists and the lay public need to understand about this process. In the project’s fieldwork and analysis, we will work to identify and articulate the visible and invisible characteristics of active, ongoing infrastructuring (Star and Bowker, 2002).

Thomas Hughes has provided a foundational study of the building of large networks (Hughes, 1983), which has led to a number of studies of large-scale sociotechnical systems, and particularly infrastructures, including power-lines, railroads, and telecommunications (Mayntz, 1988; Summerton, 1994; Coutard, 1999). It is the complexity, and the involvement of heterogeneous elements, which characterize these systems. Hughes and others have identified such phenomena as ‘reverse salients’, where in a complex large-scale system, particular elements will fall behind in developmental or technological means, resulting in a slow-down of the whole. Our work with GEON, one of this research’s foci, over the last year and a half¹ has already been possible to identify such reverse salients at the organisation level: while much energy is spent on the development of computer applications for the geosciences, it has been difficult to direct energy to the co-ordination of these services. In short while CI has the nominal goal to develop large-scale interoperable systems, the diversity of its membership tends to launch them in varying directions with little central organizational co-ordination. The scope of CI projects is ambitiously large – GEON stretches across a half-score of disciplines under the umbrella term geoscience, but also must include pedagogical access and public data distribution – the end product must *already* appeal to a vast constituency of loosely defined players. This project will undertake the basic research needed to develop tools for the formative evaluation of cyberinfrastructure projects by locating sources of reverse salients; and will develop a curricular arm to train information managers and computer scientists to strategize judiciously about combinations of organizational, social and technical reverse salients (it is the ability to choose how to intervene across a complex sociotechnical system which characterizes successful CI implementation).

Bowker and Star (1999) have studied the building of large-scale information infrastructures, such as the International Classification of Disease (ICD), the classification of nursing work or the International Classification of Viruses. They have demonstrated that social, political and organizational values play a key role in the development of frameworks for large scale data sharing. The ICD is not simply a static and uncontested taxonomy of disease; it is an information infrastructure which includes many national and international organisations, medical bodies, government and non-government institutions – each with their own agendas. Currently in its tenth edition, each revision necessitates a careful negotiation between scientists and statisticians, medical practitioners and government agencies. In short, classification is a particularly politicised affair. Phenomena such as the creation of ‘boundary objects’ and ‘boundary

¹ Please note that Bowker and Baker have been collaborating with LTER and GEON for some time, and Baker has been closely tied to the development of Ocean Informatics, thus our research plan is of formalization and scaling-up rather than a simply a future project.

infrastructures? –enable the sharing of resources between disparate constituents and permit the co-ordination of work across boundaries of expertise, thus facilitating communication (Star and Griesemer, 1989). While it appears that LTER speaks to a well defined constituency of ecoscientists, in reality practitioners in each discipline may have very different research goals, understandings of the earth, or resources for disseminating results to wider audiences. Shared organizational co-ordination mechanisms, such as boundary objects, can serve to alleviate these difficulties.

Just as we find that there are highly disparate research interests and goals in the three communities to be studied, which render the emerging CI highly difficult to describe and design, Peter Galison in his studies of technical communities within physics (Galison, 1997) has found that particular communications settings are necessary for facilitating work and research. Research is often most fruitful at the intersection of expertises (note the rise of ‘bio-informatics’) but physics is classically divided between theoreticians and experimentalists, but also between the often ignored group of instrument makers. Galison notes that in some of its most productive moments in history, physics has had strong ‘trading zones’ -- locations and languages for communication between expert groups (cf Duncker, 1998). Without the development of trading zones, communication remained internal to a sub-group, say the theoreticians, but scientific productivity was lessened. Similarly in cyberinfrastructure work, the development of trading zones is significant for cross boundaries between IT and domain sciences and education, but also within the domain sciences which must come up with a holistic vision of what kind of infrastructure will serve the community. As Berg and Timmermans have shown for medical decision making (Berg, 1997; Timmermans and Berg, 2003), the task of creating interoperable data is partly technical and partly about disciplining (or, to put it more benignly, educating) the user community. Insights such as reverse salients, boundary objects, and trading zones –

together with input from the burgeoning social scientific literature on standardization and education - are crucial for the successful development of scientific CI.

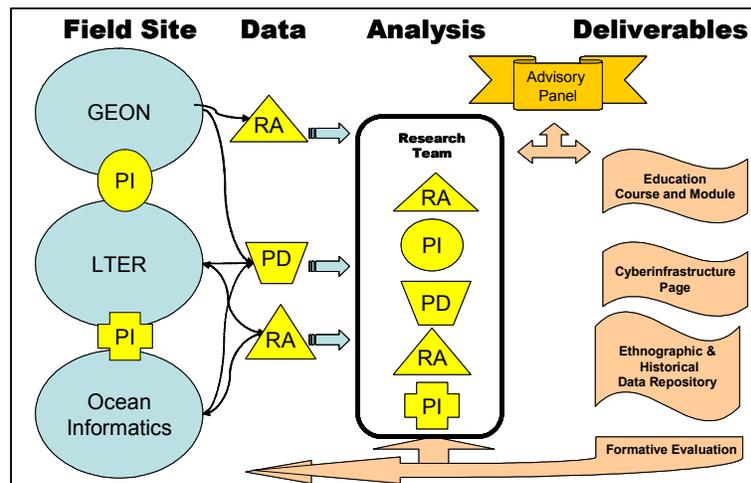


Figure 2. Workflow diagram summarizing communities (on the left), project participants (Bowker as liaison with GEON and LTER; Baker as stakeholder with LTER and Ocean Informatics; two research assistants and a post doctoral student), and deliverables (education course and module, CI Page, and Project Repository).

Three Sites, Three Approaches to Data Interoperability

Table 1, above, provides an overview of the three communities in this comparative study. Although of differing sizes and maturities, the communities are similar in their focus on disciplinary co-ordination and recognition of the need for data interoperability. Figure 2 illustrates how the proposed project participants relate to the three partnered communities and to the work detailed in the work plan below.

Geosciences Network (GEON)

GEON is one of a number of projects that together comprise the NSF Advanced Cyberinfrastructure Program. Spanning a larger constituency than any other contemporary cyberinfrastructure project, GEON is an attempt to provide the newest IT technology to the geosciences. It is also an attempt to democratize

scientific knowledge by making it both an educational resource and accessible for multidisciplinary research:

Geosciences Network (GEON) - In collaboration with the U.S. Geological Survey and the Geological Survey of Canada, this effort involving researchers from 13 universities is building digital libraries of high-quality geological information and integrated software tools for data access, analysis, modeling, and visualization. GEON will be a national resource for researchers, students, teachers, and the public. (Blue Book' Supplement to the President's Budget in Networking and Information Technology Research and Development program)

Technically, this involves integrating the management of distributed computing resources, high-speed communications, and IM capabilities as well as access to remote instruments and visualization devices in a single coherent entity (Berman, 2001). The ultimate goal of the project is to provide for the development of a more holistic picture of earth processes than is possible with the current information infrastructure. The NSF-funded GEON proposal constitutes a superb example of the kind of work that is going on in many fields of science and human endeavor to best use the multiple data sources and high data flows that characterize all of modern science and most of the important policy work that has a scientific basis (one need only think of the prospective role of the Global Biodiversity Information Facility (www.gbif.net; cf Bowker, 2000) informing international biodiversity policy and the role of heterogeneous data in world climate modeling (Edwards, 1999).

This rich project faces two major challenges. First, is the technical work of designing information systems which can incorporate vastly heterogeneous datasets. This involves the design of a web-based architecture for providing flexible views of these datasets for particular researchers; and developing stable metadata standards for the multiple communities that make up the field of geoscience. Second, closely related, is the challenge of negotiating this design and these standards with the multiple communities to be incorporated into the new infrastructure. These challenges are inextricably bound up with each other. This becomes clear if we examine a bold claim in the grant proposal that projects like GEON will lead: "to an intellectual transformation of the entire science". It is surely true that the current generations of information technology are leading to just such a transformation.

Long-Term Ecological Research (LTER)

The Long-Term Ecological Research Network (<http://lternet.edu>), initiated in 1980 by the National Science Foundation as co-operating teams of researchers where each team is a group of interdisciplinary research participants studying a particular ecological biome. LTER is currently a federation of twenty-four research sites involving more than 1200 scientists and students conducting field research locally as well as cross-site investigations (Hobbie et al, 2003; Hobbie, 2003). From its inception, LTER placed an emphasis on working with technology and preserving data for the long-term. Data management is incorporated in several distinct ways: as part of each site's individual site vision, as part of the federation of sites meeting annually as an information management community-of-practice, and as part of the centralized LTER Network Office data management team.

The LTER community has matured through its first 'decade of long-term science' and a second 'decade of large-scale research', entering now into a third decade, the 'decade of synthesis'. Site coordination develops through twice a year annual meeting discussions between representatives from each site, the Network Office and partners. Once every three to four years an All Scientists Meeting is held to bring together the diversity of LTER participants including scientists, information managers, students, field technicians, and partners. Data interoperability is currently occurring along three strands: individual site development of data, metadata, and catalogues; partnered development of the Ecological Metadata Language (EML) and tools; and actions initiated by the LTER IM committee and a Network Information System Committee of information managers and scientists working together on data management issues. After years of discussion groups on metadata and shared projects, at their annual meeting in 2000 in Madison, Wisconsin, the LTER IM Committee voted unanimously in support of EML. Sites are working jointly in interpreting what 'support EML' means and in incorporating EML into

their local work practices. The extent of effort involved has been chronically unarticulated, underestimated, and under funded.

Ocean Informatics

Ocean Informatics is an emerging project that is building community with members and selected partners within a single oceanographic institutional department. Participants are co-designers in creating an information environment with user-centered infrastructure and are from information and science studies as well as the oceanographic domain. The vision for the department, which currently supports a data center including a variety of independent data systems, is to create an information environment including a design studio that takes advantage of a variety of updated technologies and techniques to create a forum for learning, tool sharing, and participatory design. Co-design of metadata across three specific data sets is used in co-ordination with communication forums including co-ordinated seminar, workshop, and report series. This project draws in part on the tools of science studies in general and ethnographic methodologies in particular. A fundamental understanding is that heterogeneity often exists for good reasons. User-centered, participative infrastructure definition and design is under development. Qualitative techniques for observation and interview are used to capture work practice methods and design requirements that include both the technical and social aspects of a work environment that supports scientific practices.

Empirical Concentrations

One analytic approach to the study of interoperability across communities is to articulate each community's strategy for interoperability as distinct empirical concentrations. Despite the necessarily arbitrary nature of a stepwise description, such a categorization is useful for its ability to focus comparative inquiry. The following section focuses on the loci of observation that we have selected for each of the subject projects of our research. We have also provided explanations for why each locus is relevant to a holistic and comparative understanding of interoperability strategies.

GEON

GEON has chosen ontologies as its primary framework to ensure interoperability. In short, ontologies are machine executable translations between category systems (such as database schemas), permitting interoperability while requiring minimum standardization (Musen 1992, Gruber 1993). GEON's ontology work is spread between three necessary goals for successful ontology development, where each goal can be temporally ordered for the logical success of the next step: *1-education and enrolling, 2- building the ontology, 3- community outreach and uptake*. The GEON community is currently at steps 1 and 2.

Since within GEON the ontology approach has been chosen primarily by the IT team and a handful of geo-science lead scientists, it has been necessary to 'sell' the ontology approach to a broader community of geo-scientists participating in GEON. The novelty and the technical details of ontologies as an approach to interoperability require educational strategies by IT for the geo-scientists to ensure technical understanding and competence; in short, it is crucial to enroll the geo-science community by convincing them that ontology development is worth the investment of their time. This is the first empirical concentration within GEON; it is a continuing process which occurs at GEON all-hands meetings, the PI meetings, and general encounters of GEON members.

The second empirical concentration is the actual practice of developing ontologies, this occurs at GEON 'concept-space' workshops. These workshops are for the production of scientific workflows and ontologies, they are one of the points of greatest interaction between IT and domain sciences. The concept space workshops are held irregularly. They are two to four day meetings in which IT experts and geo-science domain specialists meet to discuss the production of ontologies and workflows. These

workshops are an exemplary case for IT/domain interactions; although IT and geo-scientists come from radically divergent disciplinary grounds, the successful production of an ontology is predicated on reaching understanding between participants. The majority of time in these meetings is spent formalizing the domain knowledge of geo-scientists and explaining these formalizations to IT experts. These IT computer scientists, in turn, will encode the geo-scientific knowledge into ontologies. Science requires specific conceptual agreement, there is no room for error or misunderstanding. This is doubly so in the case of ontology building, since the results become embedded within computer software applications; misunderstandings at the conceptual level of an ontology can have serious repercussions for the utility of tools produced by GEON and the data made available to users. Our research will focus on discussions between IT and domain participants, techniques used in order to understand each other, and will identify bottlenecks or systematic misunderstandings.

The final empirical concentration is the dissemination and uptake of ontologies. Because GEON is meant to serve a large constituency -- the earth-science community -- part of its mandate includes the successful deployment of interoperability beyond the PI team within GEON. This will involve educating a broader community of geo-scientists as to what ontologies are available, how to use them, and what their advantages are. Without this deployment of ontologies to a broader community, it will be difficult to represent GEON as a successful case of CI development.

LTER

The LTER community's attention to metadata can be represented by four empirical concentrations that have occurred over time: *1-discussion of metadata within information management committee; 2-design of an Ecological Metadata Language (EML) under the auspices of a professional organization; 3-partnership formation for pursuing funding to support development of EML and associated tools; 4-implementation of EML at individual sites and community evaluation of implementation.* The LTER community is currently at step 4.

Local science is a primary driver for each site's work, but embedded within each site is a data manager who attends an annual LTER information IM Committee meeting. This meeting is central to creating the LTER IM community-of-practice which has long experience in consensus decision making and strong traditions of respect for heterogeneity. Our research team will focus on the management committee's primary methods of communication: conference calls, email list serves, development partners with www.ecoinformatics.org, meeting reports and a community newsletter.

The drive for a standardized Ecological Metadata language within LTER arose from within the community discussions. Information managers came to agreement in 1992 on defining a minimum set of metadata elements across the sites and making use of one site's distribution standard (CPR now SGS). Metadata discussions continued annually within the LTER IM committee while work within an Ecological Society of America Standing Committee with some LTER participants developed in parallel, culminating in publication of an EML standard in the journal *Ecological Applications* (Michener et al, 1997). This first publication of an ecological standard will serve as a point of comparison with subsequent manifestations of the standard. In 1999 a partnership formed between the Network Office, the National Center for Ecological Synthesis and Analysis, and the San Diego Supercomputer Center that received NSF funding for a three year joint project to develop a standards based open architecture knowledge network, to develop metadata tools that work with this standard, and to disseminate products via a graduate level program and the individual LTER sites. An analysis of the proposal will serve as a basis for understanding this seminal moment in securing funding, and its organizational result. This initial understanding will be supplemented by interviews with stakeholders and augmented by subsequent documentation about the developed EML standard.

LTER site implementation workshops were held in 2002 due to impetus from the funded partnership and an LTER site (CAP) that received a synergistic grant focused on metadata tools and use. Implementation at the individual sites is a complex process. A recent 34 question site survey shows seven

of the 21 sites responded that they had implemented EML as of 2003 but only three of these having all of their datasets with at least basic EML. The survey provides a summary across the sites and represents an opportune moment of reflection upon a decade of metadata work. The survey marks a divide between the *strategy* for interoperability success and the practical implementation of interoperability. It is a cross-site integrative artifact that will add insight to our analysis of the implementation of standards and community engagement.

Ocean Informatics

The Ocean Informatics team's work is presented in five empirical concentrations: *1-team building; 2-shared project understandings constructed through use of boundary objects and participatory design; 3-IM community building; 4- community metadata design; 5- metadata implementation for targeted collections*. The Ocean Informatics community is currently at step 2.

The Ocean Informatics approach draws from lessons learned from the LTER community experiences and from an earlier project of Bowker and Baker (Baker et al, 2002), specifically recognizing the value of communities-of-practice and of participatory design in developing infrastructure to elicit local knowledge, negotiate understandings, and build community prior to proposing a joint project. A recent transition within the Integrative Oceanography Division (IOD) created an opportune time to initiate discussions on designing shared computational resources. Over the period of a year, an Ocean Informatics team assembled with explicit interests in development of long-term CI, creation of an open source environment, and interdisciplinary design of an information environment. Participatory design methods were used to engage all stakeholders equally; boundary object concepts brought to the foreground the amount of time spent on construction of shared tasks. These approaches elicited tacit information important to negotiated understandings. Preparation of several proposals was viewed as a process of working with boundary objects and of articulating infrastructure needs. The second empirical concentration entailed a joint proposal writing period aimed at obtaining funds to support continued as well as expanded discussions. Support has been requested both for individuals associated with the key IOD data collections as well as for those not-yet-identified but associated with the local institutional community. A division mission statement, a facilities description and the proposals will be the subject of analysis efforts and will be supplemented by the team's joint publications.

Empirical concentrations 3 & 4 will proceed as funding permits. The team plans community building at the division and the department level through seminars, workshops, and through an electronic report series focusing on informatics in general and metadata, interoperability, and design in particular. Informatics community engagement is seen to require an emphasis on development of professional participation mechanisms, joint exploration of collaboration tools, and evaluation strategy prototypes. Avenues for community participation in metadata design will be identified taking into account existing community scenarios and future coordination with national standards. Knowledge-making at a local level is viewed as a critical element of support for those projects that champion the establishment of standards within a discipline. Information on this part of the communities work will be gathered by attending project meetings and interviews with stakeholders.

The work is grounded by the three long-term data collections that represent a microcosm of formidable data issues. The need for development of a division and a project web site provides a visible forum and is recognized as a mechanism that prompts valuable project discussion and definition. The project aims, to develop a common infrastructuring language within the division and to contribute to a shared institutional awareness of metadata and standards, will be analyzed through their representation on the respective web sites.

Methods

At their inceptions cyberinfrastructures do not exist as anything more than an ad hoc social network of experts collected from diverse fields with a heterogeneous array of datasets. The work of beginning a

CI is that of building a common series of goals and expectations across domains as well as between domain scientists and computer scientists, then creating a functional division of labor, and securing an organizational structure to ensure long-term accountability. Thus, at the beginning cyberinfrastructures are integrally technological artefacts and social and organizational endeavours.

For example, in the last year and a half, since its inception, this team of social informatics researchers has already been working on the GEON project. GEON held its kick-off meeting in November of 2002, an assembly which primarily served to introduce the IT team, and their planned technologies, to geo-scientists. Many of these geo-scientists were only loosely familiar with each other, coming from diverse sub-disciplines. Approximately one year later, in December 2003 GEON began to purchase its hardware stacks to build its physical infrastructure, and is currently deploying this GRID (a system for the distribution of data, resources and tools, see Buyya 2002). In short, cyberinfrastructures exist initially as a social network, require many years to become a formal organization, and even upon the deployment of a physical infrastructure remains primarily bound together by human work. Social informatics research is methodologically tailored to study these social networks and emergent organizational forms: primary methods include ethnography, interviewing and content analysis (Denzin et al., 2003).

A discussion of the research practices we have been employing for the last sixteen months in the study of GEON and since 2002 in the study of LTER will serve as a surrogate for future methods, although in the later years of the grant there will be a shift towards web-based and quantitative evaluation, depending on the development of the infrastructures themselves. Another shift from our recent research will be a change of scale and diversity of locations: as the GEON, LTER and Ocean Informatics projects mature as organizations, accruing participation on the part of the domain-science community, it will be important to diversify the work of the research team, as well as matching the variety of geographic locations in which these cyberinfrastructures occur.

For the past years social informatics researchers have been attending and participating in community meetings, conferences, workgroups, e-mail discussions, and informal get-togethers. This form of research is known as participatory design and participant observation or action research as it involves both the collection of data and evaluative feedback to the participants (Blomberg et al, 1993; Schuler and Namioka, 1993). Throughout our work with these communities, we have provided formative evaluation, both officially in presentations, but also informally in discussion of organizational structure and inter-community tensions (Engestrom, 1987, 2000). The role of the social scientist differs somewhat among the communities studied as indicated in Table 1.

Data collected is ethnographic, interview and textual. We will use ethnographic and participatory design methodologies (Schuler and Namioka, 1993; Star, 1999). All meetings and interviews are tape recorded, with consent, and partially transcribed. All textual material, including presentations and slides, articles, schedules and so on, are collected and archived. A systematic archival system will be devised in order to produce an ethnographic and historical data/artifact repository. Data analysis will be conducted using computer assisted qualitative data analysis software. Somewhat labor intensive from the point of view of data entry and maintenance, computer assisted systems do not replace familiarity with the data, nor do they do the analysis for the researcher. What they do promise is to assist in the management of large data sets by providing a flexible formal structure of storing, coding and retrieving notes, interviews and memos. The software application NVivo by QSR International will be a valuable tool for working with the rich qualitative datasets collected. Computer assisted data analysis will be complemented by periodic group sessions between project researchers, this serves to ensure that all researchers know what is occurring across the communities and to produce synergistic understanding through collective analytic work. Finally, regular formative evaluation provided to the GEON, LTER and Ocean Informatics communities also generates valuable feedback as to the validity of our own research.

Interviews will be conducted throughout the data collection period of the grant, and will focus on the project manager's and PI's, but also on key IT players and geo-, ocean- and eco- scientists within the subject communities. Approximately 25 interviews per field site will be collected and transcribed yearly.

These interviews will serve to understand the subjective success of cyberinfrastructure deployment in the geo-science community; since we cannot rely solely on technological success to evaluate interoperability, we must also focus on the general enrollment of the larger science communities as participating constituencies of cyberinfrastructures.

Technical literature produced both from the IT and domain-science components will be collected, analyzed and archived. Technical literature can serve as a surrogate for the success of a CI. A successful CI will result both in publications by users, but also collaborations across traditional disciplinary boundaries. The qualitative analysis of literature will also be able to bring-forth changes in research fronts of users: e.g. is Ocean Informatics having an impact not just on how research is conducted but on what objects are researched? Thus technical literature serves both as a marker of the success of cyberinfrastructure in usage but also of cyberinfrastructure in effect.

Crucial to any social informatics endeavor is the willingness of participants to make themselves available to researchers and provide access to relevant sites. GEON, LTER and Ocean Informatics as a whole, have provided excellent access, and data collection has been greatly facilitated by this. Graduate student David Ribes is an accepted GEON community member; as a stakeholder in both LTER and Ocean Informatics, Baker provides access and insider perspectives for these projects.

Workplan

Table 2 (included below) presents details of the proposed project three year work plan with columns representing Years 1 through 3 and the rows dividing the project work into three sections: technical, analysis, and products. Technical infrastructure includes establishing some basic computational center services such as web and storage followed by consideration of group information flow with individual interfaces with central services a priority as artefact aggregation becomes a priority in Year2 and visualization a priority in Year3 with a group display unit in Year3. Collaboration tools include purchase and individual use of the software application NVivo for management, mark-up and indexing of digital transcriptions of interviews in Year1 followed by collaborative use with the Merge software and web site summarization in subsequent years. Participant tools such as videoconferencing and content management systems will be deployed and reported on in subsequent years. The ethnographic fieldwork for the three communities is central to this project and will proceed in parallel through work practice observations, meeting attendance and semi-structured interviews followed by follow-up interviews and cross-site analyses in Years2 and 3. Analysis is regarded as an ongoing process emphasizing themes of data interoperability and cyberinfrastructure attribute definition but will also include focus on community formation and science work (Strauss, 1994).

Evaluation is considered both a topic of research in order to explore formative approaches as well as a task to carry out. Evaluation will occur at three levels: with the analysis group and a focus on fieldwork and tools via weekly subgroup and biweekly group meetings; with a project advisory panel (see Table 3) through both meetings and conference calls scheduled twice yearly and additionally as needed; with the community through presentations at the partnered community forums on the themes of data interoperability in Year1, Community formation in Year2, and CI and Sociotechnical Bridges in Year3. Panel member responsibilities include participating as appropriate in panel conference calls and attending a short meeting at UCSD each year. The panel is an important sounding board that provides guidance, both grounding and broadening our interdisciplinary work. Participants were chosen to add insights with respect to national centers, history of science, digital libraries, ethnographic methods, communities, and interdisciplinary studies. Samples of panel acceptances & comments are included as supplementary materials.

Table 2. Work Plan

Elements	Plans and Products Year 1 (2004-2005)	Year 2 (2005-2006)	Year 3 (2006-2007)
TECHNICAL INTRASTRUCTURE & FIELDWORK			
Computational Support Analysis Software	services establishment: web and storage	participant interfaces w/central services	visualization & display unit
Collaboration Tools Analysis Software Participant Tools	purchase; individual use purchase; individual use	collaborative use collaborative use	collaborative use; web site presentation collaborative use; web site presentation
Fieldwork LTER GEON Ocean Informatics	observe work practice; attend meetings & interviews observe work practice; attend meetings & interviews observe work practice; attend meetings & interviews	attend meetings; follow-up interviews & observation attend meetings; follow-up interviews & observation attend meetings; follow-up interviews & observation	analyse & synthesize across sites analyse & synthesize across sites analyse & synthesize across sites
ANALYSIS THEMES			
Data Interoperability	define for each fieldwork area	synthesize data across sites; plan follow-up	synthesize results; paper for <i>The Information Society</i>
Cyberinfrastructure Attributes	explore ramifications as community intervention	engage intra-community dialogue Paper for CSCW: An International Journal	engage cross-community dialogue; web site articulation Paper for <i>Science, Technology and Human Values</i>
Community Formation	explore: organizational protocols, boundary objects, routine communications, emergent communications, traditions/stories, emergent events, social protocols	engage intra-community dialogue	engage cross-community dialogue; update and synthesis Papers for each community of CI scientists
Science Work	identify for each community	identify modes of indexing interoperability by community	refine & synthesis White Paper on data communities & cyberinfrastructure Paper for <i>Social Studies of Science</i>
Evaluation with analysis group (fieldwork; tools) with panel (project) with community	weekly subgroup; biweekly group meetings; conf calls twice yearly & as needed domain forum presentation theme: Data Interoperability	weekly subgroup; biweekly group meetings; conf calls twice yearly & as needed domain forum presentation theme: Community Formation	weekly subgroup; biweekly group meetings; conf calls twice yearly & as needed domain forum presentation: theme: Cyberinfrastructure & Sociotechnical Bridges
PRODUCTS			
Internal: Ethnographic Materials	gather materials: original, synthetic/news archive web site	classify and organize	design access for project evaluate design & use
Public: Cyberinfrastructure (CI) Page	establish web services CI Page: design	gather content CI Page: prototype evaluate for project community	synthesize material across sites CI Page: monitor use evaluate for extended community
Education Outreach Undergraduate classroom Community Secondary classroom	gather CI material for PhD course	pilot course Module: multiple-domain graduate module create secondary level presentation material	teach course with improvements Module: Masters/PhD syllabi in cyberinfrastructure Module: multiple-domain secondary school module

Table 3. Advisory Panel Participants

Participant	Organization	Role	Expertise
Fran Berman	UCSD	Director, SDSC	national center, computer science
Naomi Oreskes	UCSD	Chair, Science Studies	history of science, geoscience
Brian Schottlaender	UCSD	University Librarian/Digital Repositories	digital libraries
Leigh Star	UCSD	Professor, Communication	ethnographic methods
Maria Vernet	UCSD	SIO/LTER, Phytoplankton Component	oceanographic communities
Gabriele Wienhausen	UCSD	Provost, 6th College	interdisciplinary studies
Chip Bruce	UIUC	Professor, Library/Information Science	Inquiry Page
Theme Representative (rotating)			

Publication in relevant domain community journals is indicated in the Workplan (Table 2) by the bolded ‘Paper’, ‘CI Page’ and ‘Modules’ notation. Additional categories of products range from internal to informal public to more formal education materials. A variety of ethnographic materials will be collected and archived in Year1. Discussions will be organized to consider cross-community organization and classification in Year2 so that web access to a subset of indexed material can be designed and evaluated in Year3. The concept of a ‘Cyberinfrastructure Page’ modeled after Chip Bruce's Inquiry Page provides a framework for design and deployment of a public web service. Gathering the content for this page and monitoring its use will be activities to which all participants contribute. More formal education plans include gathering material for a UCSD graduate course in Cyberinfrastructure. The course composed of a multiple-domain undergraduate module will be piloted and then improved. In Year2, discussions regarding synthesis of materials for the secondary school level will be initiated in collaboration with the Palmer Schoolyard LTER efforts. B.Simmons, a high school oceanography teacher and Palmer LTER education liaison, has created a flexible educational framework of lessons and modules anchored conceptually by the pedagogical elements of inquiry-based science and assessment-sensitive design. A multiple-domain CI lesson will be created within this framework in Year3.

Results from Previous Studies

Bowker has extensively studied the development of information infrastructures over time. His first book was an analysis of the way in which a company was able to act as ‘information broker’ between different oil companies (Bowker, 1994). The book *Sorting Things Out: Classification and its Consequences* (Bowker and Star, 1999) brought together ethnographic analysis, interview data and archival work to analyze the development of the International Classification of Diseases as a central infrastructure to the development of global epidemiological and public health work. This project analyzed the difficult and detailed organizational work which goes into creating a robust infrastructure for multiple scientific and policy organizations. As a part of the President’s Committee of Advisors on Science and Technology to produce a report ‘Teaming with Life’ (PCAST, 1988; www.nbio.gov/about/pubs/twl.pdf) on the state of biodiversity knowledge and strategies for a workable international information infrastructure. He has since worked with the OECD’s Committee on Science and Technology Policy to produce a report (<http://dataaccess.ucsd.edu>) analyzing the suite of organizational and information infrastructural changes that should accompany any policy on access to publicly funded research data (Arzberger et al, 2004) with the Science Committee of the National Biological Information Infrastructure (<http://www.nbio.gov>). His research work for the last twelve years has been characterized by his exploration of ways to best bring social science insights into the work of developing information infrastructures. This work extends several NSF-funded projects over the past five years:

- Professional Development Grant, NSF, \$60,000 to study biodiversity informatics (1998-1999);
- KDI: Can Knowledge be Distributed? NSF 9980182 (1999-2002). \$1.4 million to analyze interdisciplinary communication amongst environmental hydrologists at the NCSA, University of Illinois (one of 6 PI’s)

Bowker has been working on building a methodological and empirical base for studying the development of information infrastructures including a paper which won the AIWORC-Bell Atlantic Best Paper Award in the Business Theme Area: 'Modeling Distributed Knowledge Processes In Next Generation Multidisciplinary Alliances' (Bowker, 2000). Bowker's new book, Memory Practices in the Sciences, to be published by MIT Press, discusses the relationship between information media and working science over the past two hundred years.

Baker is an LTER information manager and CoPI on grants focusing on the Antarctic ecosystem, information management, collaborative science, and education outreach, whose interest in information flow and infrastructure design emerges from research experiences in bio-optical oceanography and data management. She has worked with the Palmer LTER team for a decade (Smith et al, 1995; Baker, 1996, 1998), with the LTER IM community as an IM Executive Committee member for six years involved in the Network Information System (Baker et al, 2000; Brunt et al, 2002) and for a number of years contributing to education efforts through development of information management community forums (Baker and Brunt, 1999, Porter et al, 2002; Baker et al, 2000). As the Palmer Station LTER Education and Outreach liaison since 1998, she is involved in designing research-education interfaces and has worked with a wide range of informal education programs including as panel member of the NSF program 'Teacher's Experiencing the Arctic and Antarctic' since 1999 (Baker, 2001). Development of a Palmer LTER education framework for secondary school scientific materials is a part of an ongoing LTER research site-education partnership that actively incorporates assessment (Baker et al, 1999; Baker and Simmons, 2003).

NSF/DBI grant "Palmer LTER Site Information Management: Interfaces for Science and Education Interoperability" (Baker-SIO/Vildibill-SDSC) initiated work on institutional archives in collaboration with the LTER Network Office, the San Diego Supercomputer Semantics Interest Group, and the European Center for Nuclear Research (CERN). Synergistic to this effort, a site description directory was redesigned and migrated to the LTER Network Office (Baker et al, 2002). Experience handling digital artifacts including bibliographic citations while coordinating across both institutional and individual collections, was key to developing a multi-perspective approach to design of an application and to understanding critical infrastructural elements. Results have been discussed and presented within the Digital Library community (Baker et al, 2003; Gold et al, 2002). Ongoing work with the Ocean Informatics project incorporates participatory design methodologies in local infrastructure development as summarized in a recent invited talk at an e-Science Requirements meeting (Baker, 2004).

A 2002 NSF/BDEI/DGO grant "Designing an Infrastructure for Heterogeneity of Ecosystem Data, Collaborators and Organizations" (<http://pal.lternet.edu/projects/02dgo>; Baker et al, 2002) supported an interdisciplinary collaboration between Bowker and Baker. Post Graduate Researcher Helena Karasti, from the Information Processing and Systems Department at Oulu University, joined the project bringing an expertise in information systems, ethnography, and participatory design. A series of publications summarize this work: bridging social science and environmental science through design (Melendez and Baker, 2002; Baker et al, 2002), information management (Karasti and Baker, 2004; Bowker and Baker, 2004), collaborative science (Karasti et al, 2003), metadata (Karasti et al, 2002), and community dialogue (Baker and Bowker, 2001; Baker and Karasti, 2003). An organizational ethnography of the LTER is in preparation, serving to 'continue the data dialogue' with the LTER Information Management Community. In addition, six Master's students at Oulu University are working with Karasti, using the rich body of LTER materials collected.

As a post-doctoral researcher, Millerand will bring to the table unique set of skills - both theoretical and practical - which will be of great use for GEON. She has training in communication, and in particular in the area of 'cognitive technologies'. In her doctoral thesis, she has studied in-depth e-mail uses by faculty from different disciplines and analyzed the various appropriation strategies developed by users. She has showed how this new communication tool is incorporated differently depending on work practices. She has developed an approach where computer tools are understood as 'cognitive

technologies' i.e. as tools for thinking with which participate to the general production of meaning. The results of her study show signs of scientific practices-in-evolution which affects the conditions of scientific knowledge production. Millerand also has training and professional experience in computer science. As an ergonomics specialist in the Montreal Research Center for Computer Science for two years, she has gained very relevant professional skills in the design and development of software solutions, and in making working together computer experts (computer scientists) and non-expert communities (users). She will be invaluable in the design of the 'Cyberinfrastructure Page'.

Conclusion and Expected Project Significance

It is stated in the program call that: "HSD aims to increase our collective ability to anticipate the complex consequences of change; to better understand the dynamics of human and social behavior at all levels, including that of the human mind; to better understand the cognitive and social structures that create and define change; and to help people and organizations better manage profound or rapid change. Accomplishing these goals requires a comprehensive, multidisciplinary approach across the sciences...". In this proposal, we have drawn together an interdisciplinary team, with expertise in ethnography, information science, social studies of science and technology, ocean science, computer supported cooperative work and the history of the geosciences. These skills are needed to address a central vector of change in society today: the advent of a new medium for the storage and communication of data as significant as the invention of the printing press (Bowker forth coming, 2005). The dream of the eighteenth century encyclopedists was to create a universal library of knowledge freely accessible to all. We stand poised to realize a version of that dream through the information revolution. This is particularly important at a period when our relationship as a species with our environment is very delicate and must be mediated through intensive use of scientific data (Serres, 1990). However, we have a set of organizational and social forms which are tied to the older modes of enquiry and data storage. Universities typically still favor research by individuals in their reward structures; and they favor disciplinary research over interdisciplinary research in tenuring. Further, we have not yet explored new modes of enquiry enabling us to really think with the new technology rather than trying to replicate the old technology in a new medium – a factor that Paul David calls our attention to as causing the 'productivity paradox' (declining productivity with the adoption of new technology over a twenty year period) for both factories in 1880s introducing the electric dynamo and companies in the 1960s introducing computers (David, 1985).

We cannot address all of the issues raised at this critical juncture simultaneously. This project is one of a suite of projects aiming to build a rigorous understanding of the social and organizational factors involved in the development of cyberinfrastructure, and to inform the design process, the education of cyberinfrastructure professionals, and the public understanding of issues of values in the development of new information infrastructure. As Lev Manovich has remarked, the database is the central symbolic artifact of our times (Manovich, 2001). By concentrating on the strategies for data interoperability, we have for the purposes of this proposal centered on a fundamental issue in the development of a cyberinfrastructure which is destined to change the ways we think, learn and act in the world.

Intellectual Need for the Proposed Activity

The emergent cyberinfrastructure will fundamentally change our ways of understanding the world and managing its complexity. It is crucial at this foundational moment to develop sustained, critical understanding of the social and organizational dimensions of this work. Three interdependent aspects of change agency are developed within the scope of this proposal: exploration of a range of approaches to data interoperability that will impact both the currently studied and future scientific data-sharing communities; contribution to the shaping and definition of cyberinfrastructure that emerges from such a comparative study approach – to be incorporated into educational materials; and a deeper understanding

through joint engagement and articulation of the strengths, tensions, and mechanisms of interdisciplinary work itself.

This proposal provides support to bring together a comparative study from three ongoing longitudinal studies of scientific communities. The interdisciplinary team proposing this work has collaborated productively over the past three years. Co-location at UCSD with strategic components of the partner communities combined with institutional interest and support that is incorporated in an advisory panel, further enhances collective opportunities. This work brings together environmental scientists, social scientists and technologists to work proactively on issues of data use and integration, taking into account the diverse roles of technology.

Studying cyberinfrastructure in all its forms requires such an interdisciplinary effort, which will set the stage for breakthroughs in understanding the complex technological and social processes inherent in data use both for collaborative scientific work and for science policy. The recognition, articulation, and discussion of differing strategic approaches within multiple communities simultaneously will make a valuable contribution to the development of interdisciplinary sensitivities and skills required for the growing number of globally scaled projects. Working actively to understand and to characterize the complex sociotechnical dynamics of shared scientific cyberinfrastructure, its deployment and ramifications, is critical to our understanding of how data and technology are integrated into any society of multiple constituencies. Through this study, we develop a grounded understanding of the organizational complexity in producing shared scientific cyberinfrastructure, and specific understandings of the cost and benefits of three interoperability approaches: metadata standards, ontologies and community-driven approaches.

Broader Impact Resulting from the Proposed Activity

The development of scientific cyberinfrastructure is vital for this country's future economic prosperity and for its ability to respond to key policy issues with scientific and technical dimensions. Cyberinfrastructure is a large-scale contemporary investment; this study will help inform the decisions which today are determining future structural outcomes.

The project will facilitate understandings at the level of science policy of the organizational and social dimensions about building shared infrastructure. We will produce a policy white paper on data communities and scientific cyberinfrastructure. It will suggest guidelines for the ongoing formative evaluation of infrastructuring activities. In conjunction with development, there is a need to develop educational programs – both inreach and outreach – which sensitize domain scientists, computer scientists and science policy workers to social and organizational issues.

We propose a three-pronged educational project that creates an interdependent set of broader impacts by– providing ways of *reading* cyberinfrastructure at the secondary school level (a key tool for the informed citizen of the twenty-first century); ways of *creating* it by developing a curricular kernel for merging science studies and information science to train the next generation of cyberinfrastructure information managers; and finally ways of *interacting* with it by developing a Cyberinfrastructure Page for the CI design community. Bowker will develop, as a centerpiece to a local push for a Masters level program in cyberinfrastructure, a graduate course in the development of cyberinfrastructure. Baker will develop a secondary school lesson module on cyberinfrastructure within the pedagogical framework of an ongoing scientific research-education partnership. The course and module will be incorporated into a 'Cyberinfrastructure Page' website which will enable us to share results first within our partner communities and then across communities. This will provide the kernel for a resource to be developed as a site for researchers and practitioners in the emergent field of scientific cyberinfrastructure to share findings and best practices and to engage in collective problem-solving.

References

- Arzberger, P., P. Schroeder, A. Beaulieu, G.C. Bowker, K. Casey, L. Laaksonen, D. Moorman, P. Uhlir, and P. Wouters, An International Framework to Promote Access to Data, *Science*, 303, 1777-1778, 2004.
- Baker, K.S., Development of Palmer Long-Term Ecological Research Information Management, in *Proceedings of Eco-Infoma Workshop, Global Networks for Environmental Information, Lake Buena Vista, FL*, pp. 725-730, Environmental Research Institute of Michigan (ERIM), Ann Arbor, MI, 1996.
- Baker, K.S., Palmer LTER information management, *Data and information management in the ecological sciences: a resource guide (Proceedings of workshop, held at University of New Mexico, Albuquerque, NM, 8-9 August, 1997)*, 105-110, 1998.
- Baker, K.S., Palmer outreach: a diverse group with common goals, in *The LTER Network Newsletter*, edited by P. Sprott, pp. 8, Long-Term Ecological Network Office, University of New Mexico, Albuquerque, NM, 2001.
- Baker, K.S., Ecological Design: An Interdisciplinary, Interactive Participation Process in an Information Environment, *Proceedings of the Workshop on Requirements Capture for Collaboration in e-Science, 14-15 January 2004, Edinburgh*, 5-7, 2004.
- Baker, K.S., B.J. Benson, J. Brunt, N. Gardiner, D.L. Henshaw, E. Melendez-Colom, J.H. Porter, P. McCartney, and D. Steigerwald, LTER Information Management: Paradigm Shift or Paradigm Stretch?, LTER All Scientists Meeting, August 2000, Snowbird, Utah, 2000.
- Baker, K.S., B.J. Benson, D.L. Henshaw, D. Blodgett, J.H. Porter, and S.G. Stafford, Evolution of a multisite network information system: the LTER information management paradigm, *BioScience*, 50 (11), 963-978, 2000.
- Baker, K.S., G. Bowker, and H. Karasti, Designing an infrastructure for heterogeneity in ecosystem data, collaborators, and organizations, *Proceedings of the Second National Conference on Digital Government Research, 20-22 May 2002, Los Angeles, CA*, 141-144, 2002.
- Baker, K.S., and G.C. Bowker, Designing an infrastructure for heterogeneity of ecosystem data, collaborators, and organizations, in *The LTER Network Newsletter*, edited by P. Sprott, pp. 9, Long-Term Ecological Network Office, University of New Mexico, Albuquerque, NM, 2001.
- Baker, K.S., J. Brunt, and D. Blankman, Organizational informatics: a research network site description directory design (accepted), *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*, 7, 355-360, 2002.
- Baker, K.S., and J.W. Brunt, Information Management Newsletter of Long-Term Ecological Research Network, in *LTER Databits*, Long-Term Ecological Research Network Office, University of New Mexico, Albuquerque, NM, 1999.
- Baker, K.S., A.K. Gold, and F. Sudholt, FLOW: co-constructing low barrier repository infrastructure in support of heterogeneous knowledge collection(s), in *Proceeding of the 2003 Joint Conference on Digital Libraries(JCDL'03), 27 May - 31 June, Houston, Texas*, pp. 397, IEEE Computer Society, New Brunswick, NJ, 2003.
- Baker, K.S., A.K. Gold, F. Sudholt, H. Karasti, and G.C. Bowker, Palmer LTER: Information Flow and Management, in *LTER All Scientists Meeting 2003, Embarking on a Decade of Synthesis*, Seattle, Washington, 2003.

- Baker, K.S., and H. Karasti, A Whirlwind Tour for LTER IMers: ACM, CSCW, ECSCW, IRIS, in *LTER Databits Newsletter*, Long-Term Ecological Network Office, University of New Mexico, Albuquerque, New Mexico, 2003.
- Baker, K.S., D.S. Rawls, S. Bell, B. Dawson, M.A. Wallace, and W. Winn, Proceedings of Palmer LTER Education Outreach Forum, held at National Center for Ecosystem Analysis and Synthesis (NCEAS), Santa Barbara, CA, on 25-28 July 99, in *SIO Ref. 99-14*, pp. 55, University of California, San Diego, Scripps Institution of Oceanography, La Jolla, CA, 1999.
- Baker, K.S., and B.E. Simmons, Palmer Long-Term Ecological Research Education-by-Design; Looking at Sea Ice Data (poster and online lesson), Americal Geophysical Union; Science Education Resource Center, Cutting Edge Workshop, Seattle, 2003.
- Berg, M., *Rationalizing Medical Work - Decision Support Techniques and Medical Problems*, MIT Press, Cambridge, MA, 1997.
- Berman, F., The Human Side of Cyberinfrastructure, in *EnVision*, 2001.
- Blomberg, J., J.GIagomi, A.Mosher, and P.Swenton-Wall, Ethnographic Field Methods and Their Relation to Design, in *Participatory Design: Principles and Practices*, edited by D. Schuler, and A. Namioka, pp. 123-155, Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.
- Bowker, G., Modelling Distributed Knowledge Processes in Next Generation Multidisciplinary Alliances, in *Academia-Industry Research Conference on Working Challenges*, edited by A.C.-B. Atlantic, 2000.
- Bowker, G., and K. Baker, Information Ecology: Open System Environment for Data, Memories and Knowing (accepted), *Journal of Intelligent Information Systems. BDEI Special Series*, 2004.
- Bowker, G., W.Turner, S.L.Star, and L.Gasser., *Social Science, Technical Systems and Cooperative Work: Beyond the Great Divide*, Lawrence Erlbaum, Mahwah, NJ, 1997.
- Bowker, G.C., *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920-1940*, The MIT Press, Cambridge, London, 1994.
- Bowker, G.C., Biodiversity Datadiversity, *Social Studies of Science*, 30 (5), 643-684, 2000.
- Bowker, G.C., *Memory Practices in the Sciences*, MIT Press, Cambridge, MA, forthcoming, 2005.
- Bowker, G.C., and S.L. Star, *Sorting things out : classification and its consequences*, 377 pp., MIT Press, Cambridge, Mass., 1999.
- Brown, J.S., and P. Duguid, *The social life of information*, x, 320 pp., Harvard Business School Press, Boston, Mass., 2000.
- Brunt, J., P. McCartney, K.S. Baker, and S. Stafford, The future of ecoinformatics, *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL, 7*, 367-372, 2002.
- Buyya, R., Economic-based Distributed Resource Management and Scheduling for Grid Computing, Ph.D. thesis, Monash University, Melbourne, 2002.
- Coutard, O., *The Governance of Large Technical Systems*, Routledge, 1999.
- David, P., Computer and Dynamo: The Modern Productivity Paradox in a Not-Too-Distant Mirror, Center for Economic Policy Research, Stanford University, Stanford, 1985.
- Denzin, N.K., and Y.S. Lincoln, *Strategies of Qualitative Inquiry*, Sage, Thousand Oaks, Calif., 2003.

- Duncker-Gassen, E., *Multidisciplinary Research at the University of Twente*, Twente University Press, Enschede, 1998.
- Edwards, P., Global Climate Science, Uncertainty and Politics: Data-Laden Models, Model-Filtered Data, *Science as Culture*, 8 (4), 437-72, 1999.
- Engestrom, Y., *Learning by Expanding*, 371 pp., Orienta-Konsultit Oy, Helsinki, 1987.
- Engestrom, Y., H. Ahonen, and J. Virkkunen, Knowledge Management -- the second generation: Creating Competencies within and between work communities in the Competence Laboratory, in *Knowledge Management and Virtual Organizations*, edited by Y. Malhotra, Iea Ge, Hershey Group Publishing, 2000.
- Futrell, J., and AC-ERE, Environmental CyberInfrastructure: Tools for the Study of Complex Environmental Systems (<http://www.nsf.gov/ere>), 2003.
- Galison, P.L., *Image and logic : a material culture of microphysics*, xxv, 955 pp., University of Chicago Press, Chicago, 1997.
- Gold, A.K., K.S. Baker, K. Baldridge, and J.Y. Le Meur, Building flow: federating libraries on the web, *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries 2002, 14-18 July, 2002, Portland, OR*, 287-288, 2002.
- Gruber, T.R., A Translation Approach to Portale Ontology Specification, *Knowledge Acquisition*, 5, 199-220, 1993.
- Hobbie, J.E., Scientific accomplishments of the long-term ecological research program: an introduction, in *BioScience*, pp. 17-20, 2003.
- Hobbie, J.E., S.R. Carpenter, N.B. Grimm, J.R. Gosz, and T.R. Seastedt, The US long-term ecological research program, in *BioScience*, pp. 21-32, 2003.
- Hughes, T.P., *Networks of Power: Electrification in Western Society, 1880-1930*, Johns Hopkins University Press, Baltimore, 1983.
- Karasti, H., and K.S. Baker, Infrastructuring for the long-term: ecological informationmanagement, in *Proceedings of the Hawai'i International Conference on SystemSciences (HICSS) 2004, 5-8 January, Big Island, Hawaii*, IEEE, New Brunswick, NJ, 2004.
- Karasti, H., K.S. Baker, and G.C. Bowker, Ecological storytelling and collaborative scientific activities, *SIGGROUP Bulletin*, 23 (2), 29-30, 2002.
- Karasti, H., K.S. Baker, and G.C. Bowker, ECSCW 2003: Proceedings of the Computer Supported Scientific Collaboration Workshop, Eighth European Conference on Computer Supported Cooperative Work, Helsinki, Finland, 14 September 2003, pp. 55, University of Oulu, Finland, 2003.
- Kling, R., What is Social Informatics and Why Does it Matter? (<http://www.dlib.org/dlib/january99/kling/01kling.html>), *D-Lib Magazine*, 1-23, 1999.
- Kling, R., and W. Scacchi, The web of computing: computer technology as social organization, *Advances in Computers*, 21, 1-89, 1982.
- Lave, J., *Cognition in Practice: Mind, Mathematics, and Culture in Everyday Life*, Cambridge University Press, Cambridge, UK, 1988.
- Manovich, L., *The language of new media*, xxxix, 354 pp., MIT Press, Cambridge, Mass., 2001.
- Mayntz, R., and T.P. Hughes, *The Development of Large Technical Systems*, Westview, 1988.
- Melendez-Colum, E.C., and K.S. Baker, Common information management framework: in practice, *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*, 7, 385-389, 2002.

- Michener, W.K., J.W. Brunt, J.J. Helly, T.B. Kirchner, and S.G. Stafford, NonGeospatial Metadata for the Ecological Sciences, *Ecological Applications*, 1, 330-342, 1997.
- Musen, M.A., Dimensions of knowledge sharing and reuse, *Computers and Biomedical Research*, 25, 435-467, 1992.
- NRC (National Research Council), How People Learn: Brain, Mind, Experience, and School, edited by J. Bransford, A. Brown, and R. Cocking, pp. 374, National Academy Press, 2000.
- NSF-AC-ERE, Advisory Committee for Environmental Research and Education, in *Complex Environmental Systems: Synthesis for Earth, Life, and Society in the 21st Century*, edited by S. Pfirman, 2003.
- NSF-AP, Revolutionizing Science and Engineering Through Cyberinfrastructure(http://www.communitytechnology.org/nsf_ci_report), in *Report of the NSF blue-ribbon Advisory Panel on Cyberinfrastructure*, edited by D. Atkins, et. al, Messerschmitt, Ostriker, Wright, 2003.
- PCAST, Teaming with Life: Investing in Science to Understand and Use America's Living Capital., in *President's Committee of Advisors, Science and Technology, Panel on Biodiversity and Ecosystems.*, 1988.
- Porter, J.H., K.S. Baker, and S. Stafford, The Ecoinformatics Challenge: Meeting ecological information needs for the site, network, and community, in *The 6th World Multiconference on Systemics, Cybernetics and Informatics*, Orlando, FL, 2002.
- Schuler, D., and A.Namioka, *Participatory Design: Principles and Practices*, 319 pp., Lawrence Erlbaum Associates, New Jersey, 1993.
- Serres, M., *Le Contrat Naturel*, F. Bourin, Paris, 1990.
- Smith, R.C., K.S. Baker, W.R. Fraser, E.E. Hofmann, D.M. Karl, J.M. Klinck, L.B. Quetin, B.B. Prezelin, R.M. Ross, W.Z. Trivelpiece, and M. Vernet, The Palmer LTER: A long-term ecological research program at Palmer Station, Antarctica, *Oceanography*, 8 (3), 77-86, 1995.
- Star, S.L., The Ethnography of Infrastructure, *American Behavioral Scientist*, 43 (3), 277-391, 1999.
- Star, S.L., and G.C. Bowker, How to Infrastructure, in *The Handbook of New Media*, edited by L.A. Lievrouw, and S.L. Livingstone, pp. 151-162, SAGE Publications, London, 2002.
- Star, S.L., and J.R. Griesemer, Institutional Ecology, "Translations," and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39, *Social Studies of Science*, 19, 387-420, 1989.
- Star, S.L., and K. Ruhleder, Steps toward an ecology of infrastructure: design, access for large information systems, *Information System Research*, 7 (1), 111-134, 1996.
- Strauss, A.a.J.C., Grounded Theory Methodology, in *Handbook of Qualitative Research*, edited by N.D.a.Y.S. Lincoln, pp. 273-285, SAGE, Thousand Oaks, CA, 1994.
- Summerton, J., *Changing Large Technical Systems*, Westview, 1994.
- Timmermans, S., and M. Berg, *The gold standard : the challenge of evidence-based medicine and standardization in health care*, ix, 269 pp., Temple University Press, Philadelphia, 2003.