Title
The Long-Term Information Management Trajectory: Working to Support Data, Science and Technology

Authors
Baker, Karen S
Karasti, Helena

Publication Date
2004-06-01

The Long-Term Information Management Trajectory: Working to Support Data, Science and Technology

Karen S. Baker
Integrative Oceanography Division, Scripps Institution of Oceanography
University of California, San Diego, USA

and

Helena Karasti
Department of Information Processing Science
University of Oulu, Finland

This is an organizational ethnography of some facets of the Long-Term Ecological Research (LTER) program that opens a small window on a complex, heterogeneous community carrying out scientific research collaboratively.

Scripps Institution of Oceanography Technical Report
June 2004

Preface

This report gathers together the views of a community, juxtaposing them according to the understandings of our research team. As an organizational ethnography of the Long-Term Ecological Research (LTER) community, it provides a small window into a complex, heterogeneous community carrying out research collaboratively. The report is necessarily incomplete as the community is large and the story evolving. This study was embedded within and stretched across multiple domains: ecological and social informatics, ethnographic and science studies, and ecological and information sciences.

The work of environmental science is to gain an understanding of an object of study within its context and to inform colleagues about the results of study. The LTER adds an additional dimension, the concept of long-term data care. Collaborative science adds yet another research dimension when the focus is on participative design and co-development: the need to dialogue with those involved with the study. Establishing and maintaining the connection between our research and the LTER community members has been an important, ongoing and time-consuming challenge. We write this report as part of a continuing dialogue with LTER participants including the information management community-of-practice embedded within the larger LTER community.

We began this study in 2001 with a broad definition of the LTER community, allowing for multiple objects of study as we considered the data, the individuals, the communities-of-practice, and the LTER community as well as their relationships. We adopted a user-centered ethnographic approach recognizing that collaborative and information habits are often not well informed by theory. Yet, it is the context, whether in the form of formal research results or informal community stories, that enables the creation of cumulative knowledge especially at the interface of the physical and social worlds. Use of qualitative research methods has produced a rich body of ethnographic materials and insights into the community as well as into how to work with and within a lively, ongoing, on-growing community.

The goal is to offer not remedies but representations of a community's multiple voices. Our aim is to prompt reflection regarding the work of information management that LTER participants may choose mindfully in creating information systems and interfaces. There are a myriad of critical elements in the developing practices of collaborative science and in the developing roles associated with digital information environments. As in the past, when participants engaged at the practical level are both actively engaged and enabled, the strengths of diversity in a federated network are tapped. Taking an optimist's view, the promise of the 21st century will begin to unfold when we begin to do our work, our everyday information practices, from new perspectives.

Karen Baker
June 2004

Over the past twenty years, new forms of distributed collective work practice have been developed which are having a huge impact on business, government and science. With the use of an array of new networking technologies, it is now possible to 'explode' large organizations into distributed groups which operate seamlessly as a virtual organization. There are two fundamental shifts going on here: the development of the new networking technologies and the development of new databasing technologies. There has been much academic attention to ways in which people work together or not within these new frameworks once they are in place. However, remarkably little attention has been paid to the work of the information manager, who sits between the computer scientist and the domain scientist and makes the myriad adjustments and designs the many procedures that turn a suite of loosely coupled applications and heterogeneous databases into a working infrastructure.

In this project, we have attempted to elucidate the key role that information managers play in this process, building our analysis from interview data and from observation of work practices. What we have to offer back to the IM community is our perception of the contours of their work. It is a commonplace in the literature on professionalization (Schön 1983; Abbott 1988) that key to the development of a nascent profession is the development of a discourse about the scope, reach and techniques of their activities. We hope that this report can serve as one seed for this important process.

Geof Bowker
June 2004

Abbott, A. (1988). *The System of Professions: An Essay on the Division of Expert Labor*. Chicago, University of Chicago Press.
Schön, D. A. (1983). The reflective practitioner : how professionals think in action. New York, Basic Books.

Table of Contents

The Long-Term Information Management Trajectory: Working to Support Data, Science and Technology

Karen S. Baker and Helena Karasti

Abstract

The everyday work practices of Long-Term Ecological Research (LTER) information managers are complex and contribute in multiple ways to scientific research. Our work aims to go beyond the formal image of data management and make visible some of the aspects involved in the day-to-day work that support the LTER program vision. Our focus is on work practices and the information management defined by multiple relations and tensions and structured to adapt to change processes. Major issues in the work of information managers are described in this report through their own voices.

The everyday work practices of an information manager encompass technical and social issues related to data management and information processing. The work requires juggling a multitude of tasks and timeframes as well as sustaining multiple roles and memberships. Such complexity and fluidity may be conceptualized as an information management trajectory. This trajectory accommodates change and yet is often invisible due to lack of articulation.

Three intertwined aspects of information management, namely support for science, data, and technology, are discussed. The tensions in data environments and information management work created by balancing these divergent elements are described.

Keywords: information management, Long Term Ecological Research, LTER, work practice, community of practice, metadata, trajectory, change, technology, mediation, invisible work, articulation.

Foreword

With this report we wish to continue a dialogue with the Long-Term Ecological Research program information managers. We have intentionally gathered together quotes capturing the multiple views from interview materials. Our aim is to present the diversity of participants with their differently situated views. Further, we acknowledge the interpretive nature of the transcription process (Polland, 1995). Some excerpts may sound strong or even extreme as they are separated from their original context and others have become contextualized through juxtaposition with the thoughts of fellow community members. The excerpts may also sound colloquial, to the eyes and ears of one not familiar with ethnographic methods and qualitative interview research. With the intention of retaining the residuals of everyday experience, we have not engaged in ‘cleaning them up’ (except for omitting some non-relevant parts). One interview was lost due to an unfortunate technological mishap. We are sorry for not being able to provide this important voice.

We have done our best to remove clues leading to individual identities. For the purposes of this dialogue, it is less important who the individuals are than what they represent. In order to provide context, participant roles are indicated: graduate student (GS), information manager (IM), scientific principal investigator (PI), research associate (RA), and technician (T).

The quotes are numbered to facilitate cross-referencing. Together, these many voices can begin to describe the heterogeneous aspects and relations of long-term ecological information management today.

1 Introduction

The initial vision of the Long-Term Ecological Research (LTER) program included the concept of data management (Callahan, 1984) as a part of each site's science plan. Today there is an integration of information management (IM) embedded within the research program at each of twenty-four LTER ecological research sites and the LTER Network Office (LNO). This approach creates a forum within which the information manager role is co-constructed by participants (e.g. scientists and technicians, research assistants and students) simultaneously within the local site and within the information manager community of practice as data, technology, and science community needs are addressed.

> "So from the very beginning, [with] the LTER projects and how they were conceptualized in NSF, IM was a central part of it because they realized that in order to have these datasets that were going to be going for decades, there had to be a system where the dataset would be curated and taken care of, and resources needed to be provided for that, and a set of expectations were created for what would be the components of IM. Those expectations have evolved over time." (IM) 1.

> "I think information management is viewed as an important and critical part of the LTER and in large measure because, well, one, the expectation is mandated by NSF but, two, because we are creating a legacy of data. A legacy of data doesn't have any sense if you don't have information management in a way to get at it and to know what is there, and how it has been done, the metadata and everything that goes with it." (PI) 2.

The LTER program was established in 1980 (Franklin et al., 1990; Magnusen, 1990; Swanson and Sparks, 1990; Hobbie, 2003; Hobbie et al., 2003). Site scientists and data managers at the first LTER sites began to discuss connectivity and technology and to gather data into centralized, local repositories in the early 1980's. This period of data collection is the initial 'decade of long-term research' (Brunt et al., 1990). In the 1990's, new ways were developing to address public accessibility, web presence and cross-site connectivity (Michener et al., 1994, 1998). Concurrently, LTER site science was broadening from a focus on local ecological studies to regional ecosystems as well as to cross-ecosystem and cross-domain themes (Brown, 1994; Strebel et al., 1998). This time of data description, assembly, and coordination has been described retrospectively as the 'decade of large-scale research'. The third decade of the LTER, the 'decade of synthesis' (LTER, 2002), will require an even greater information management focus on data organization and integration as well as new approaches to communications, knowledge sharing, and collaborative science.

The larger culture of global technology sets the framework within which developments occur for the LTER information management community. Personal computers and digital networks had burst upon scene when the LTER began in the 1980's. Applications such as WYSIWYG word processors and spreadsheets brought new mental models and standards into everyday use in science work. The second decade of the LTER coincided with the appearance in the 1990's of hypertext, the world wide web, browsers, search engines and applications supporting document availability and exchange. As we enter the third decade of the LTER program, the global digital community faces new questions of security and ethics while two important developments are in progress: the semantic web and metadata concepts. A semantic web vision has been articulated (Berners-Lee et al., 2001) although the enabling, ubiquitous infrastructure, tools and applications are under research and development while the concepts and the particulars of metadata are unfolding.

The LTER data management practices in support of these three decades of LTER science are summarized in the literature (Michener, 1986; Michener et al., 1994; EcoInforma 1996, DIMES 1998, and SCI2002: see Appendix 8.3; Michener and Brunt, 2000). Although domain literatures focus separately on technology, computer science and environmental science, the work of information management is occasionally reported formally (e.g. Stafford et al. 1994, Strebel et al., 1994, Olson and McCord, 1998; Olson et al., 1996, 1999) and more frequently through more informal avenues such as community presentations or newsletters (http://lternet.edu/databits). Comments on information management along with thoughts on organizational practices are sometimes provided in hindsight as 'lessons learned' (Stonebraker, 1994; Thorley and Trathan, 1994, Strebel et al., 1998; Baker et al., 2000; Benson and Olson, 2002). Within LTER, a variety of the individual site approaches to information management have been published (Baker 1996, Benson 1996, Briggs and Su 1994, Ingersoll et al. 1997, Porter 1996, Spycher et al. 1996, Stafford et al., 2002, Veen et al. 1994). In addition, there are summaries of the LTER network federation itself (Brunt and Nottrott, 1996; Brunt, 1998; Baker et al., 2000, 2002; Brunt et al., 2002; Karasti and Baker, 2004). Despite differences between individual IM systems, there are aspects of the IM work that are common across sites.

Continuing from the earlier LTER work, this report aims to bring information management practices to the foreground, and to make some taken-for-granted and/or unarticulated aspects both visible and accountable. Rather than focus on information systems and databases, we bring attention to how the work is actually done. The information process within the scientific realm, presented by a number of LTER practitioners below, is influenced by technological, social, and cultural elements in addition to scientific objectives and the data itself.

1.1 Our approach to studying work as everyday practice

Our research is motivated by a profound interest in *everyday work practices* and *information ecologies*. Through empirical study we approach *work practices* as a form of unfolding activity in actual communities that is concrete, complexly socially organized and technologically mediated (Blomberg et al. 1993, Karasti 2001). We employ a research approach that has roots in anthropology and sociology and has increasingly become used also in technology settings and workplace studies (e.g. Plowman et al. 1995). *Ethnography* is committed to study the activities of people in their *natural settings*. This derives from a belief that particular behaviors can only be understood in the everyday context in which they occur and they have to be fitted into the larger whole (*holism*). Ethnography involves understanding the work from the *members' point-of-view*, that is how people organize their behavior and make sense of the world around them. Based on fieldwork, a *descriptive* understanding of the work practice is developed. (Blomberg et al. 1993)

By *information ecologies* we mean an analytic approach with associated epistemological assumption which pertains to the very observations and scrutiny of data, practices, collaborations and infrastructures denoting concrete everyday work practices and situated knowledges. Such an analytic approach requires an interest in the mundane, even 'boring' or 'singularly unexciting' things (Star 1999) as well as seeing knowledge and meaning as socially constructed within ongoing communities of practice (Blomberg et al. 1993) instead of, for instance, relying on a typical distinction in division of labor into 'routine/knowledge work' (Blomberg et al. 1996). Differing views exist with respect to day-to-day practice and knowledge work (Figure 1; Blomberg et al. 1996). A 'hierarchical' view of organizations assumes that work carried out in lower organizational levels is dominated by routine work, parts of which may be automated through technology. Within this framework, knowledge work is regarded as 'higher order' and to be performed by the top levels' personnel. In contrast, an 'integrated view' tips this hierarchical triangle, recognizing that all jobs intertwine both routine and knowledge work components.

This epistemological starting point allows for analyzing and accounting for the less obvious in information managers' work. Their work is often dominated by unplanned interdependent events that are more the exception than the norm so that skill and judgment are called for habitually. However, the knowledge component is not captured by traditional academic merit systems:

> "We [information managers] don't do things that are in the metrics that the PI community value. We don't write multi-million dollar grants. We don't publish a bazillion papers every year. We are too busy getting the work out the door. So based on the metric that most of the traditional scientific community uses, we are pretty invisible." (IM) 3.

Terms such as complexity and tacitness are used as researchers explore and describe the 'invisible' elements of everyday practice (Star & Strauss 1999, Suchman 1995). In describing the sociology of the invisible, Star (1991) notes that "articulation work is work that gets things back 'on track' in the face of the unexpected, and modifies action to accommodate unanticipated contingencies".

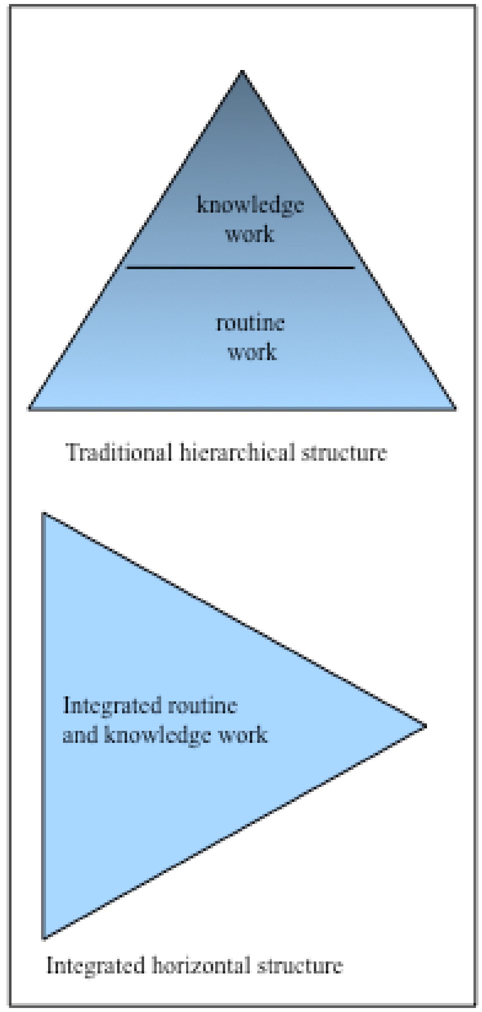


Figure 1. Two differing work practice models (after Blomberg et al, 1993).

1.2 Our research process and methods

The composition of our research team, both outsider and insider LTER investigators with varied backgrounds, allows for the integration of ethnographic (Becker, 1998) and action research approaches (Greenbaum and Kyng, 1991; Schuler and Namioka 1993, Karasti, 2001). Our work is predicated on the importance of empirical work including continuous collaboration and dialogue between the fieldworker and the other research team members, and permitting the investigators to be part of the community in order to develop as holistic an understanding as possible of work and collaboration.

Ethnographic interest in understanding human activity in everyday settings calls for *participant observation* (Atkinson and Hammersley, 1994) because people often cannot readily articulate some aspects of their work that are so familiar to them as to be unremarkable. Participant observation at one particular LTER location - that accommodated a number of members of one distributed site - continued for the entire year. This was supplemented by observation of both virtual and co-located meetings of the site. Shorter, more focused visits to a number of other LTER sites comprising observation of selected local work practices together with interviews with site personnel provided for understanding local contingencies and practicalities, as well as differences between the sites. To gain insight to network level activities, participant observation was also carried out in various meetings organized during a period of nine months, including the network's science, coordinating committee and executive committee meetings as well as information managers' committee and executive committee meetings. Extensive debriefings and discussions in the research team guided the fieldwork.

Interviews with LTER participants and associates were carried out throughout the year. The research team prepared themes in advance but sessions were conducted in an open manner that allowed for serendipitous and in situ topics and elaborations (Holstein and Gubrium, 1995). Interviewees were selected to cover a variety of LTER sites and all major roles, e.g. scientists, information managers, research assistants, field personnel, technicians, administrators, and graduate students together with LTER associates, and to provide – as much as possible – for diversity in views and perspectives. Discussions in the research team provided for more informed choices in the selection of interviewees.

The fieldwork resulted in over 50 transcribed interviews that averaged approximately 2 hours in length, 10 notebooks of field notes, a variety of paper-based and electronic materials, and digital pictures. Qualitative analyses, both individual and collaborative, of interview materials (Silverman, 1993, 1997) started during year 2002. The emerging themes include information management, boundaries, tensions, change, scientific collaboration, and stories (Karasti and Baker, 2004; Karasti et al., 2003).

Gradually, as the fieldwork progressed, more elements of action research (Reason and Bradbury 2001, Coghlan & Brannick 2001) were intertwined. We created opportunities for participant reflection by sharing with the community our observations and developing understandings. We have presented initial findings for comment and discussion, for instance, in information managers' executive and annual committee meetings, via postings on our project website, and in writing in the community newsletters. This report on the long-term information management trajectory is a mechanism to continue the dialogue. In comparison to a more traditional stance of (participatory) system design (e.g. Bødker et al. 1993), our interventions have been moderate (e.g. Heath 1997). While extensive ethnographic fieldwork was carried out during the year 2002, analyses by the research team and dialogue with the LTER community continue.

2 Major issues in the information management trajectory

A conceptual grounding for the development of information work over time is provided by the notion of a *trajectory*. *Trajectory* is an analytic term coined to refer to the *total organization of work* done over the unfolding course of events, taking account of the impact on those involved with that work and its organization (Strauss et al. 1985). With the trajectory concept our aim is to capture the complexities and fluidity of information management.

An LTER information management *trajectory* includes both site level work and network level activities. Site work involves providing research service and support, managing data, as well as designing and obtaining support for changing technologies; network level activities involve sharing and learning within a community of practice, understanding heterogeneous development strategies, and planning to accommodate program changes. Information managers bring together the various, often-incompatible aspects of their everyday work practice by juggling a multitude of tasks and timeframes as well as by sustaining multiple roles and memberships.

2.1 Managing Data

Long-term studies today require data to be recorded consistently, documented adequately, archived digitally, and accessed electronically. In addition to the variety of general information management challenges, two data issues are particular to LTER research: data heterogeneity and long-term timeframes. LTER teams, working with an array of heterogeneous data, must specifically address long-term data concerns such as data storage and reuse as well as legacy data and contextual information.

2.1.1 Heterogeneous, long-term data

Ecological data are by nature especially heterogeneous. Data sets differ both within a site and across sites. They may be of deep or wide coverage, with differing spatial and temporal extents, within single or multiple fields as well as of recent, legacy and/or external origin.

> "[S]o that is one of the things that make people nervous about studying about ecology because it is so, not just complicated, it is so spatially heterogeneous, it is very difficult to study." (IM) 4.

> "The LTER is a much different situation because it is a much more diverse heterogeneous dataset with all these different systems. So I think there is a big challenge there." (PI) 5.

> "We have a lot of varied types of datasets. Some studies may have a ton of records, a "deep database", not a lot of diversity but huge volumes (like remote sensing). In ecological data in general you get much smaller databases that cover a much wider variety, "wide databases". In general you are struggling with the diversity of different types of data, therefore generic modes of maintenance are a challenge. In genetics for example, in comparison, databases are deep but not the amount of diversity." (IM) 6.

Heterogeneous ecological data result, for instance, from differing spatial and temporal sampling, from the variety of instruments and dissimilar methods as well as the diversity of data sources.

> "LTER sites are required to submit renewals every 6 years and because each renewal has to be interesting and exciting and different, we have seen the natural progression from site research to scaling up spatially and regionally and being more involved in cross-site and new site research." (IM) 7.

> "You know, if the method changes, we have to re-document how things work, it might compromise the original data." (IM) 8.

While data heterogeneities differ between sites in the LTER network, the sites all face a similar challenge with respect to maintaining long-term data.

> "What is the real value? 50 years of stream data and then we had that flood in '96; what is the effect on flow? New questions get posed to long-term datasets that make it even more valuable (public interest). A database does really increase in value over time if it's well maintained, even though it may lose some of the historical facts, the overall value is going to increase." (IM) 9.

The ramifications of caring for the data over the long term are numerous and significant. Not only must the data be accessible for current work, it must be contextualized through its metadata if it is to be understandable over time for researchers in the future:

> "LTER is characterized by the long-term research. In that sense it is not only separated from other fields but also from other ecology projects. There is an implicit (and also explicit) need to preserve the data, the repeatability of experiments, the understandability of these datasets. We have had a 20-year rule, a 100-year rule. These milestones have been set up to us by people in talks, but the idea is that the data collected now should be understandable and the studies repeatable 20 years down the road. Some have extended that to 100 years. The 20-year rule we bring up routinely. They have to be documented and stored in a way that they can be preserved. It is difficult to some researchers who are used to the traditional way, 2-3 year projects and then it is over. Data to be comparable not only empirically but also statistically … requires a lot of documentation because so many things change." (IM) 10.

A shared understanding among LTER information managers, illustrating the need and the motivation for capture of long-term data, is summarized in a frequently referenced figure (Michener et al., 1997):

> "There is the classic Michener chart (information decay figure). General figure to show, to stimulate information management: what can you prevent? The more it [data] is described with metadata, the more is known." (IM) 11.

> "Historically whatever IM was done, was done by the actual researchers on the project and that led to the loss of a lot of data because people were not organizing their efforts for long term preservation and sharing with larger community, above how people share through the literature and publications. Numerous cases of where valuable data sets were either lost or were not sufficiently documented so that you would have confidence in the methods used…" (IM) 12.

> "They call it data decay, loss of information over time that tends to happen. It can happen remarkably fast. We have had anecdotes about little studies that were done and they attempted then to go and recreate them two years later and no one could remember how they were done. Not properly documented." (IM) 13.

> "Well one of the things you find is if there is not somebody looking over things on a day by day basis as it is coming in, you are going to lose certainly. The dataset will already start to lose value. And basically the Michener graph which is very true, I think, data you collected from that point, from the moment you collect it, time goes by you begin to forget things about it, so the value starts to go down, immediately. (chuckle) Although his graph says the most you know about the data is the time of publication, I would really take issue with his graph there and say, really when you know the most about the data is when you first look at it, if that means when you collect it, or you first pull the data logger up and you look at it, you probably know the most right then because you know whether the data logger was running or not, and functioning well, or you know the point as you are writing it down, and so you really."(IM) 14.

2.1.2 Ecological Science Context

A holistic ecosystem view developed in the 1950's along with some national and international associations (Carson, 1950; Odum, 1953; Worster, 1977). Ecological datasets were collected during the International Biological Program (IBP: 1964-1974) established as part of the International Council of Scientific Unions and the International Union of Biological Sciences and sponsored in the United States by the National Academy of Sciences and the National Academy of Engineering (Golley, 1993). The IBP preceded LTER, focusing on the immediate potential of mankind to damage earth's ecosystems and initiating data collection at or near some of today's LTER sites (Andrews, Coweeta, Jornada, Niwot Ridge, North Temperate Lakes, Short Grass Step/Central Plains Experimental Range/Pawnee, Konza Prairie and near Barrow):

> "…going way back to the IBP, there was always a legacy of databanks. It was always looked as a plus, not a drag. There were always enough researchers that were willing to use the repository and saw the advantages…" (IM) 15.

> "one of the things that the IBP was notorious about being bad at, was the long term management of data and information. And that is why when David Kingsbury and John Brooks and Tom Callahan first had the idea of an LTER, they were really focusing on making sure that someone designated to keep track of the data was identified in the very, very beginning. Because it had not been done effectively in the IBP. At the IBP, data was viewed as a resource in a rare book room. It was a volume on the shelf in the library that no one was to check out. Whereas data in the LTER was viewed as a dynamic body that was to be used and to be accessed and to be manipulated and then put back so that someone else could use. But the derivation of information and value from the data was continual." (IM) 16.

Participants in the International Biological Program recognized the need for metadata as documentation for long-term data:

> "…in the 70's there was IBP (International Biological Program), which was a precursor to LTER. They really started figuring out ways to document datasets, managing data for long term if it was going to survive." (IM) 17.

LTER information managers preserve and pass on the history of the concept, working with professional organizations such as the Ecological Society of America.

> "Major emphasis of the IM group over the 10-12 years has been on documentation. They call it now metadata, we called it then data documentation when we started. At that time we thought of metadata as machine readable documentation, but we now want all of our documentation to be machine readable." (IM) 18.

> "Documentation and metadata. It's really unlike anything that has been done in ecology, and it does preserve datasets over time. Ecological Society of America has tried to identify datasets at risk, important ones to the discipline as a whole and to get them documented. That has been based on the work done in the LTER network, as far as establishing what needs to be documented, the practices… The network has had a great influence, pushing forward a standardized approach (not standards) to collecting metadata." (IM) 19.

The community of LTER information managers developed and exchanged metadata forms, thus introducing the concept/requirement as a tool into the site data process. This process took place over some years in a variety of discussion forums. These early metadata efforts provided opportunities to inform scientists, to train information managers, and to mature local responses prior to looking beyond local data form requirements to consider interoperability with national and international metadata standards.

2.1.3 Data Access and Policy

The availability of internet access technology created new expectations for data:

> "Those expectations have evolved over time, since the emergence of the WWW and Gopher and the ability to put data online. This was a new expectation, not only were the datasets going to be managed well and documented but they were going to be made readily available to a wider community." (IM) 20.

… and also gave impetus to issues of data access and acknowledgment policy. A prompt from a US funding agency, the

National Science Foundation (NSF), regarding data sharing came at this time of growing Internet connectivity:

> "Well the data policy applies to the core data, which are datasets that the LTER say are essential to their data legacy." (PI) 21.

> "I know that LTER has the data policy and, (but I don't remember it word for word either) so its like within 2 years everything is supposed to be public,... in the case of graduate student projects that are going to take longer than that, that is an exception." (PI) 22.

Sites were challenged to meet a data policy mandate in 1994. After much discussion, individual sites created data access and data sharing policies. The LTER network data policy is posted online (http://www.lternet.edu/data/netpolicy.html).

> "The early mandate with LTER that the long term datasets will be maintained has grown into this incredibly critical part of LTER." (IM) 23.

> "Which one dataset was to be on-line first? I can [still] … remember that historic moment when ('96?, the CC-meeting?) the announcement was made (this had followed a year or two [of] discussion…internet connectivity for LTER sites, what year? '94?). We have been talking about this, let's just do it. We are the LTER; we can do something in a new way; we need to show a new path. So, they were appealing for the LTER to be progressive. So they said: every site must have one dataset online (within some timeline). I do remember the ripple of the [subsequent site] discussion, the tension….which data set should we put [online]? Which component [would contribute for our site] THE DATASET!?" (IM) 24.

> "So it [data policy] is something we are still working on, there is a lot more acceptance of the whole objective of information management than was the case say 4 or 5 years ago. There was a strong resistance because they had grown up you see in a climate or a culture that was more like individuals working within the whole group and they were, nobody asked them to do any data archiving. It was really quite voluntary, and so the whole network as I am sure you know and you have heard from others, it has been an evolutionary process going on which is fascinating to watch because it has changed, I think it has changed a lot in the last 7 or 8 years even. There was considerable resistance to the network playing any significant role, they were just something, quite a few more of my senior colleagues said well this is something NSF wants, we don't want it. That was their attitude. Now even they are saying you know this is actually a very good idea because they want to do, have more access for more broad based, cross site comparison of data, and that kind of thing." (IM) 25.

The issues of data ownership, access, and sharing arise along with the development of digital delivery capabilities.

2.1.4 Identifying the data

LTER sites collect data that contribute to an understanding of the local ecosystem and to central program themes. Each dataset involves different measurement techniques as well as particular consistency, quality and context issues. Some sites designate 'core' datasets and projects as those that will be measured and maintained over extended periods of time:

> "we have moved away from pilot projects to more these are our core projects, these are in the core monitoring projects." (PI) 26.

> "the concept of core data set (s)… those are the dataset(s) that the sites identify as the ones they are going to do over the long term." (IM) 27.

> "because it was core one time, just because it stopped, I mean, it's still a valuable data set, this isn't suddenly non-core at that point…we just decided if something was LTER or not, tried to just prioritize based on our notion of what the most important things were…But I think our notion of core is pretty broad. (IM) 28.

In addition to identifying and gathering new data at the site, there may be datasets collected in the past or associated with the site in some other way. Such datasets add to the development of a long-term perspective. If the decision is made to recover and maintain such legacy datasets, support work is required:

> "I feel like more than other sites because of the size of our legacy data sets and what we are trying to undertake in terms of the system, it's just taking more time. The corporate information, the climate and the stream flow data, those sorts of things get collected by the field crews… basically they're constantly all coming to you with coordinating things just on the corporate information. Management of this corporate information becomes another important aspect of my work." (IM) 29.

> "I mean the datasets I have that different postdocs have done, that I have on my list to get into the, the dataset, but aren't really part of the core, but we put as much as we can in." (IM) 30.

> "So everyone is really on board on this. It has become kind of a central focus of our LTER…So I think that growing sense and the actual importance of the past and the fact that we have extensive records of this going back the last hundred years for this particular property helps to strengthen the interest that our scientists have in managing long term data because of this long term perspective." (IM) 31.

Sites with pre-existing data are faced with duplicate datasets and their differences, with missing data and varying data quality as well as with ensuring data access and availability:

> "Well existing data, this is a really, I mean there are a lot of good things about it, and one is because of both environmental regulations and, you know, interest on the part of cities and other agencies too, you know, have some control over environmental quality and the quality of life ... It seems like everybody…bunches of different groups are monitoring the same things. And then there are other things for which there are no data at all. And you think how can this be? There is so much more entry going on, they haven't measured this one little thing, and then we have to go and measure it. … There is just a goldmine of data available for cities … The problem is that there are all these problems of what is the data quality, you know, how reliable are the data that we can get from various places: are they willing to release them, you know. They may have data, but they are not willing to release them. And then because you have a lot of people collecting data on the same things, what if they conflict? All that kind of thing. … there is a lot of demand for, you know, getting access to existing data." (PI) 32.

> "…social science data. Interview data, data from tax records, data from government agencies about housing purchases, housing starts, developing license permits, all those kinds of things, are used very heavily in the urban fringe project, looking at how the development is proceeding at the urban fringe, what kind of a dynamic interface that makes between the urban and non-urban environments, and what goes on within that fringe. And all of them rely heavily on a lot of kinds of data that most other ecological sites don't deal with. Census data, data on zoning, land use and development, that kind of thing." (IM) 33.

New ideas for supporting data collection may introduce change to information management practices:

> "More recently we have realized that the place where we can really capture a lot of data is through getting involved more when data is acquired, so basically in the field, when data is being acquired as opposed to after the fact, after it has been acquired, when people are processing that or whatever." (IM) 34.

2.1.5 Gathering data and metadata

The challenges to LTER information managers to get the data and requisite metadata have increased notably with the mandate to put data online, to make site datasets public:

> "… you have to be productive, but we also have to make some investments in the long term, so I think this is the biggest conflict with data management getting metadata: getting data in. People, scientists in general were sympathetic to getting a publication out. They realized that it will probably take them 20 or 30% more time if they actually really clean the data up and figure out what it is, you know, get it squirreled away and stored away properly. And some people they don't want to make that investment, other people want to

> but haven't effectively been able to do it and some people do it and you know, you got the whole range." (PI) 35.

> "writing metadata, people sharing their data and getting credit for it; I think it's the biggest challenge for data managers." (IM) 36.

Several sites have created IM protocols that outline processes and procedures. However, plenty of articulation work is required to make the protocols work in practice (cf. Strauss et al. 1985, Star and Strauss 1999, Suchman 1987). This work ranges from educating investigators and students about the importance of data management to continuous contacts and prompted reminders to investigators.

> "we started off with a protocol that was handed to us …so we adapted that, we did sort of a dry run, or a trial run in '99, in the Spring of '99, and we find our protocols thereafter. And then according to that protocol I designed a data sheet because [she] did not want to work with a computer out in the field. She thought the hand written sheets she could take and file were a little safer. So we designed a database that would take all that and …so we had the GIS database" (IM) 37.

> "I basically serve as a quality check at, well we have a written procedure, things that, when dealing with types of plants often there is some subjectivity in how one interprets how a measurement should be done. And so having historical view of if there is any type interpretation, I can come up with that explanation. So I guess the point is, a written procedure and protocol is essential, but it may also require some enhanced clarification on top of that." (IM) 38.

> "when you get new people in you have to like make them converts into thinking that having their data in our databank and on the web is something they really want to do. If they do not have the mind set that they want to share the data, it is really difficult to make them do it." (IM) 39.

> "And so the potential for loosing acquisition of the dataset …you miss that opportunity of being there at the beginning to help with the design or help them with the structure of the information of the data. Whereas a notification of research application now provides an opportunity for raising a flag to indicate 'here is a new study that is about to begin'. Gives you an opportunity to completely interact with the researcher. At the beginning make them aware of certain expectations… helping them structure the design of their experiment to provide… structured form acceptable for an information management system. And actually in many ways facilitate their work because they would be potentially unaware of the resources that we have that can help them in that." (IM) 40.

As an information manager finds it necessary to move beyond the role of 'data receiver' to adopt the role of 'data getter', a variety of strategies exist, including education, 'encouragement and prodding', 'data nagging', 'badgering'…:

> "and it's been a massive process of sort of education and badgering and you know, to get people to think that that is important" (IM) 41.

> "our lead PI is very supportive of information management. Others look at it as being a good citizen of the community. And they, you know, they always let you know, that I am being a good citizen here, when they are giving you any metadata at all. Just look, I am doing, I am being a good citizen here. Like don't forget this that I am giving you this, and it is not like something they feel like they really should do. I mean they feel like they are doing this, going way out of their way just to provide like the basic documentation for their data, and then the others you can't really get anything out of them. You pretty much, a lot of the documentation I end up pulling out of the paper. If they have written a paper on it, I can just go to the paper and get methods and things that they are not going to give me directly. So I end up tapping other resources as much as I can." (IM) 42.

> "So we require all of the PI's submit a documentation, which is just really a category where they put in you know basic information about methods with what the variable names are, where the research was, so that a form that explains as much as you can. And then along with that, with an ASCII data file that has been mostly comma delimited, with just headers on the top of the file and nothing else. And in general it has worked out well and that

> most people realize the value of it. Often there needs to be prodding to get them to find the time to do it." (IM) 43.

Information managers use multiple approaches in motivating and facilitating data and metadata submissions:

> "When they come in without the metadata you have to do an outreach there, not just send them away saying come back when you have the metadata but to offer them whatever systems and help that will help them do that. It is not that big of a task if you do it while the study goes on, but if you wait until the end it is a lot of work. You still get people that wait until the very end to do the whole thing. You need to put a carrot on a stick and to follow-up on that (in a friendly fashion continue that interaction)…. I have rubbed elbows during the years with these researchers. I know what is going on, sometimes intimately, other times superficially. Allows me to have a little bit of inside track to greased rails. Sometimes I can do something for them that is really simple so that they end up with fewer blanks to fill in. Often partly a question of cutting and pasting parts of the research notification and other documentation. Sometimes when they give me a representative sample of their data, I do the structuring for them, but ultimately it is their responsibility. Matter of leading them through the process. Sometimes quick interactions of corrections and additions. So, you have a full spectrum and information management has to be very flexible. One fixed method of doing that is not enough. Giving more than one path to give the information is needed. Web interface will work for some people but others would never do that. Structured word document will work for others (pump it through Data Junction software, to automate the metadata into the information management system). I still get people who want to do it manually. They do not have a computer, for example visiting researchers. I am very reluctant to having people leave without giving me anything. Once they are gone, there is really not any enforcement to get it." (IM) 44.

Sometimes incentives are used to interest researchers in the metadata work, for example, services are provided such as QA/QC or data handling tools (see also Porter and Callahan 1994):

> "We have some people who do a really good job of it and others do not. I think you have to sell them on some of the incentives to participate and so if you can develop an information system to do the quality assurance for them, we have the ability to quickly put the data on the web, so they don't really have to do much once it's in the system. The easier you can make it for them to put in the metadata, then I think you're going to have more success." (IM) 45.

> "it's been pretty good in that they [scientists] even go to the database to retrieve files that they had put in because it is easier for them to find it there, than go through their stuff. … Sometimes it is easier for them to go to the database to retrieve data than it is to try to find in their disks or whatever" (IM) 46.

At some sites more forceful methods have also been exercised:

> "There are some [investigators] that are really so attached to the data still … my PI had to say, either you give [the information manager] the data and fill out the documentation, the metadata, or you are not going to receive any money from me. When he said that, then they gave me the data." (IM) 47.

> "So in the past …[data sharing] wasn't happening, and so we have started every three months, we send an email out to all of the PI's with an accounting of what data is over three years old and that it will be moved. And that data is broadcast, the email is broadcast to every single Principal Investigator, and so even if you don't have data that is older than three years, you still receive that email so that everybody knows what everybody is doing … if they are not going to do what they have agreed to do, then they are going to be asked to leave the LTER, and they won't get funding." (IM) 48.

2.2 Providing services, creating rapport and facilitating alignments

Information managers provide diverse kinds of service and support to various other project team members:

> "Information management is not only dealing with data, it is also dealing with all the

investigators and helping them with a lot of issues involving technology, sometimes computers, sometimes other kinds of technology." (IM) 49.

"You know everyday life: a lot of times you help investigators with their computer problems, not necessarily data entry or data management problems, but computer administration, computer troubleshooting, hardware and software." (IM) 50.

"we still get requests by email or phone, and people drop by our office and want some kind of help (both within the group and outside)" (IM) 51.

Typical of service work (e.g. Star and Strauss 1999), this work is frequently interrupt-driven; the number of unscheduled events include the need to answer unexpected questions, to respond to immediate field or proposal needs, to handle technology change and break-down, as well as to interface with collaborators:

"there is a lot of demands … I get emails from people who download things … I am getting constant requests of questions and information, like a steady stream, hundreds of these a year, people wanting to know about a variety of things… and that is on top of everything else that you are trying to achieve … With my job setting priorities is really difficult … There are constant demands from the lead PI's. You are playing different roles at the site, you have to attend to meetings, keep up with the science as well. For me, and I think for all IMs, it is becoming more and more things are being asked for them to do. … I feel like I am getting spread too thin. For one thing, I definitely cannot concentrate on what I want." (IM) 52.

"Well my PI, for example, this week he wanted, he was writing a proposal which I was not given a copy so I wasn't quite clear what it was he was after. But he says that he really wants to look at the gradient of rainfall, precipitation amounts across the [study] Basin. Now that is really all he gave me. He was in [a name of a place in another continent] with the group that went over recently for LTER conference I think it was. So he emailed me from there with this thing like, well what is that supposed to mean? You know this is really difficult from other countries to get scientific information on precipitation ... this is a lot to do. You know there are a lot of countries that border that basin. So I, you know, I really thought it over and I emailed him back. And he wasn't clear about what it was that he wanted so it took me time and I finally made a decision. OK well look, he wants just a basic map more like a thematic map of high to low. Is there a gradient? And I made a decision to do a search on the web site, like tourism sites. What is the annual precipitation ... you know that kind of thing. And I made a map based on the tourism. This is very non-scientific. And then I found a site that there is actually a mission that measures precipitation by satellite platform, that has very nice maps across the world actually of precipitation. And so I gave him a few figures from that. But you know I look and 5 days go by again. And then I am like, oh my gosh. You know my deadline is August 15th and here I just wasted, not wasted but in terms of the main mission, those days are gone." (IM) 53.

"[It] does also mean sometimes that one would like to spend some more time on one thing and it's a little hard to do. I have been trying to write a paper for the SCI2002 meeting. Last week I was in the office 3 days and I couldn't get anything done on it at all. I wrote the introduction for it on the way down here and am hoping to make more progress on the plane back because that is at least one place where you don't get a lot of drop-in things. (IM) 54.

The ability to be able to support requires both understanding what the needs are and finding ways to provide service.

"You really have to make an effort to talk with the scientists at your site and understand what it is that they need and what it is that they want, and then try to design your system to do that…" (IM) 55.

"you have to look for ways of giving services to users." (IM) 56.

Nurturing relationships with scientists is part of the background articulation work that information managers engage in, in order to create favorable conditions for working relations between information management and science.

> "My site has really been trying to build better rapport with PI's." (IM) 57.

> "It's something that requires continual maintenance, but it doesn't have to be an onerous task to do. It can be just little things from day to day with the interactions with the researchers, that IM, information management, is maintained in a visible profile." (IM) 58.

> "We bring more emphasis on rapport than merely reporting…I think as a group we have a very nice blend of report and rapport; relation and relationship. (IM) 59.

Information Managers attend to multiple tasks and operational management of people and resources, eventually becoming one of the communication nodes for everyday coordination of collaboration for the site.

> "It's constant demands from the lead PI's … two field people down there … for me it is like coordinating with them … they go through me a lot of times, all this info coming up from there…" (IM) 60.

> "Information Management acting as a communication node between getting the science done, the scientists, and the technology. That layer is missing in our educational system today; this is what I have gained within the learning environment of LTER." (IM) 61.

> "The information manager is a hub for a lot of communications because they talk with everybody" (IM) 62.

2.3 Designing for change

The contemporary proliferation of hardware, software, and scientific instrumentation imposes new requirements for design strategies that accommodate change. Within LTER a number of strategies have evolved.

2.3.1 Changing technologies at a site level

Staying technologically informed is an important aspect of information managers' work:

> "There is need for people to remain current in technology and to do some long range planning for how the computer facilities and new technologies are going to change over time ... (IM) 63.

> "the important thing is the ability to learn new skills, having sort of an affinity for the technical types of things." (JP1/VCR-8) 64.

… yet information managers are not pure technologists:

> "…most of the data managers in the LTER sites weren't really technologists … their role was a technician who has a secondary position to try to develop some infrastructure as opposed to a computer scientist or a, you know, somebody that is really more involved in the technology that might try to apply it to ecology." (IM) 65.

The information manager plays an active role in how technology and data management concepts are introduced and sustained at a site. The information manager is often creative in promoting a sound, and perhaps advanced, information management vision in conjunction with a modest technological approach. Although staying technologically current is a driver, other factors must be considered in determining the rate at which new technology is adopted for site information management systems.

> "supporting research should drive the information technology. The researchers are looking to the information manager to guide them in that area. It is important for the information manager to be very proactive and to come up with their vision, of where the site is going and how it connects into the network level information management and to provide that plan for the site" (IM) 66.

> "new technology has changed the way people do their science. I can remember the days when people were saying I am never going to use email or people looking at some of the new technologies and thinking it was kind of flashy, like Geographic Information Systems, like the web. And then they become such an integral part of how people function and collaborate." (IM) 67.

In addition to the limitations imposed by resources, judicious decisions about technology procurement are influenced by the features of high reliability, easy maintainability, and low risk for long-term data management and science support.

> "Servers are getting better, more reliable. But as much as possible we want to maintain the individual desktops as functioning units. Even if everything else crashes down around them, they can, you know, people want to work at 2 in the morning, can still be there doing their papers" (IM) 68.

> "Maybe it is just the needing to keep all these old things going plus do all the new things, knowing that the old things also need maintenance and attention and upgrade and then doing the new things too. It's an overload of where to focus. So it's a matter of setting priorities…. So it is the juggling between what is possible and what needs to be sustainable and will be important for the future." (IM) 69.

> "And because it is a very simple measurement, it has been easy to maintain over time." (IM) 70.

The persistence of technological change prompts cautious thinking and careful balancing of options.

> "So they have this, this legacy of historical data that is, some of it is, on tape that can no longer be read because the instrument to read the tape is obsolete." (IM) 71.

> "And that's been, that experience we have had with several of our things, is that the issue isn't how you do it, it's how do you maintain it and how do you make it so that it is easily maintainable." (IM) 72.

> "given that we have the investment in [current technology], it is not so bad yet that I am willing to go back and rewrite all my interfaces." (IM) 73.

> "Thank goodness they didn't update us to Windows 2000 until after the study was done. Because for some reason Microsoft 2000 does not recognize to print flat data files, what we call ASCII files." (RA) 74.

An information manager's foremost concern in aligning existing technologies with developing technologies is to minimize disturbance of ongoing data archival and use followed by interest in optimizing long-term data use.

> "they [information managers] see the inconsistencies and non-alignments. They have the tools of technology to think about applying to some of that, and then they have the reality checks of what is really possible, having seen, ya know, doing it with the data, what is really possible" (IM) 75.

On one hand, there is the concern for having in place a data-safe, functional system for maintaining the integrity and availability of the long-term datasets. On the other hand, incorporation of new capabilities to enhance data capture, use and preservation always holds the potential for an exciting facilitation of science. Technology such as relational database software has been adopted by some sites while other sites view the investment as too high.

> "Grad student was hired to build a filter for each site to solve the problem, cool for a moment, but not sustainable: first the grad student left, so no one to maintain it, and second some sites [data structures] changed." (IM) 76.

> "A survey [of LTER sites] showed different people trying out different databases, some of the big ones not being able to be sustained, having to be changed to smaller ones, scaled down to … sustain because they could not be maintained." (IM) 77.

> Databases are relational but the overhead of making those relations, establishing them and maintaining them, is so high, we do not do it [database work] ... the software is not there yet." (IM) 78.

Ongoing changes in data taking, projects, personnel, organizations, technology and standards produce profound challenges that call for re-evaluations and modifications of IM practices and plans.

"For a long time we developed all these tools for doing generic testing of all of our databases, tools to do quality assurance of the metadata, to be able to write web pages from our database, ability to write reports of the metadata (generated metadata report), export information to the SAS, all these little utilities but they all were based on data being of ASCII format or text style format, …. They were all stored in FoxBASE database. Well, Fox database has now been bought by Microsoft and it may go away some day. You just cannot stay in the same system forever, though it is nice. This metadata content issue came up in the '97; what the standards for metadata should be paper. EML, ecological metadata language. It has elements that we have not traditionally kept track of. We have to expand the metadata, which also causes major changes to the system that we were maintaining before the changes. We have changing technology and expanding metadata content, which led to the redesign of the schema. In another five years we'll probably be asking again what do we do now? The more I hear about XML, the more I think about us having to move out from our relational databases, everything might be XML, who knows. I don't think we will abandon relational databases any time soon. Technology keeps changing..." (IM) 79.

"What we would like to do, and what we will be able to do I think in next years, is to track current and new projects quite effectively through this use of web forms. And I think we will be able to collect a lot of the metadata that we need up front…I think as part of that we will work hard to get our datasets updated more carefully than we have recently. Some are kept up to date pretty well, but others lapse and its up to us to, or up to me to, make sure that people contribute their data. Otherwise I am generally too busy to do it, on their own." (IM) 80.

Some sites facilitate the dialogue between an information manager and site participants about IM plans and visions by establishing formal mechanisms, such as an information management committee (IMC) or regular discussions in site science meetings.

"…who we've involved for the consultation, is going to be the executive committee, the data management committee, as well as selected other players within [the site]… because it is important to us that we consult the PI's." (IM) 81.

"Well it [the IMC] was started prior to my becoming, coming on board. And the data management committee is comprised of myself, and ... my colleague as well as the lead PI's … So we are trying to establish, have breadth among these people but the idea is really to, to not let the data manage us ... to actually have somebody thinking about what are proper data formats, what are proper guidelines." (IM) 82.

"Some of that involves an educational component because a lot of researchers are not aware of what really goes on in information management, and it is important for them to understand that. I spend some time just going in and meeting with the lead PI on the project and bringing him up to date about things. I recently talked to him about EML and what that was going to mean for our site, why it was important. And periodically, we have monthly science meetings, I will request that part of one be devoted to information management, and then I will just talk to the whole group about what some of the new issues are and things that I think they ought to be aware of." (IM) 83.

Site level information management strategies (e.g. Baker 1996, Benson 1996, Briggs and Su 1994, Ingersoll et al. 1997, Porter 1996, Spycher et al. 1996, Stafford et al., 2002, Veen et al. 1994) are influenced by the local environments. Information managers also address the issues in and through the LTER Information Management Committee as the following sections describe.

2.3.2 Accommodating heterogeneous and flexible development strategies within the network

On the network level the importance of local infrastructures at LTER sites and the resulting diversity of site approaches is recognized:

"a lot of the systems in terms of business and things like that are designed sort of top down. OK. And that was the approach that was taken at IBP also. It was a top down sort of a system. And that is a system where you can decree, you can say, thou shall have your forms look like this and that the good ideas will come from the top and will be executed down throughout the system. It is a much more hierarchical system. And as I said in reaction largely to IBP, LTER was designed with the significant bottom up component." (IM) 84.

"One of the things that has been critical to the success of the information managers group, during the time period I have been here, has been the recognition that there are legitimate reasons for some differences between site systems. An example that I use a lot: the Andrew's system, until very recently, was based on FoxPro databases running on PC. If you wanted to make any changes to the metadata you had to go to the PC. At our site we made the system so that all of the interfaces for changing the metadata were on the web. Which of those is the right solution? For us a computer down the hall was not going to do any good because at that point we had 26 investigators and 7 institutions in 4 states. On the flip side, at Andrews all of the investigators were down the hall; it would have been ridiculous for them to spend a lot of time and resources on developing a difficult web interface. Basically from acceptance that there are legitimate differences. …In terms of the information manager group, it is the willingness to give respect to somebody who has views different from yourself." (IM) 85.

"A lot of the bottom-up characteristics are important for LTER IM. Ability to deal with heterogeneity not by limiting it but by dealing with it." (IM) 86.

Technological heterogeneity is not only allowed, it is also seen as one of the strengths of the LTER IM network.

"Cherry picking octopus. Best practices also include looking around at what is going on within the network because you don't spend so much time looking at your own stuff that you never look at anyone elses. I learn more by looking at other LTER sites. If I see that they are doing something neat, I try to find out how they did it, the good things, the bad things. There is always some site that is out there looking for the new solutions. Cherries are the good pieces of software. There is some software that just does this job but you don't love them. They are a lot of work. There are other that just make your life easier, a lot easier. By having 24 sites, you have 24 opportunities to find good things. And the reason it has to be an octopus is they need to be connected to the center. The arms need to bring the cherry back in." (IM) 87.

"I don't feel like every site has to be at the leading edge of trying to push EML, or whatever the latest stuff is. And as far as the philosophy of us having data freely available in an easy format to download and use, I think our site has always had that philosophy." (IM) 88.

The LTER IM group has participated in developing some standards and guidelines, e.g. minimum standard installation for technology, standard installation for meteorological station, and guidelines for data policy (http://lternet.edu/data/netpolicy.html; http://intranet.lternet.edu/archives/documents/reports/Technology-reports/msi1988.html). Typically they outline a minimum set of requirements that have been developed over time and are specified at the functional level, having flexibility to accommodate individual site information management.

"When GIS and remote sensing became sort of the in-thing, the network wanted to adopt that, setup some minimum standard installation dealing with remote sensing information. So they had to be given the resources for doing that. So, NSF I think has been reasonably supportive but not just pushing some ideas, and information management is another one that we sort of tried to elevate, but also providing some resources for it." (PI) 89.

"There was something called a minimum standard installation, which was the basic amount of hardware and software that a site needed to be able to do the electronic communication, remote sensing, GIS stuff, so it was basically a workstation connected to internet and the software appropriate to GIS and remote sensing." (PI) 90.

"… [augmented] the original standard where they set out basic levels of participation of LTER sites for climate data and made general suggestions/recommendations for how data should be collected. At what height off the ground, like, you should have your wind sensors…that sort of thing, as well as how units would be reported. I mean, the earlier report, most of it was describing the level zero that would be just min-max temperature and precip. And Level 1 where you would have, you could actually generate temperature and precipitation in maybe an hourly format or something. And then there was a Level 2 which added a bunch of variables and kind of included the notion that you would have continuous data loggers and much more comprehensive complete climate stations … Everything that it is now was the rewrite of that '86 document done in '97." (IM) 91.

"At the information managers meetings we brainstormed some basic principles for IM policies. We took that to the coordinating committee and they appointed an ad-hoc committee … in some cases people who were most recalcitrant about some of these things, they made some changes, actually strengthening. Basic principles were pretty attractive: scientific data should be shared for the good of all and it should be available in a timely fashion, people should get credit when their data is used … Another thing was that we did not come up with the LTER wide information policy, we would have had endless discussion. We published guidelines for individual site policies." (IM) 92.

Information managers as a group have also started to explore and develop network-wide systems, for instance a Network Information System (NIS) including a climate database (ClimDB).

"About the time that the web was evolving and the expectation was coming to the LTER sites to put their data online, the information managers began to have a vision that has evolved into the network information system. Not only were individual sites going to be putting their data online in their own idiosyncratic ways, but we were going to design a way of accessing that data in a centralized kind of way, through one interface so you didn't have to go out and visit all 24 sites in order to get, for example, the climate data – so that's been within the last decade." (IM) 93.

"I mean, what we're trying to do with the NIS, I think, is to communicate and educate the PI's that if they are going to do an intersite effort, that they include data management as a major, a necessary, component." (IM) 94.

A lesson was learned from IBP:

"With IBP there was a big focus on big centralized databases, going out collecting data using standardized methods. And all of this data would be sent to a computing center to be entered into mainframe computers, analyzed & models would be built. It did not work out as planned. They were database professionals but they didn't understand the science aspect. Ecological databases are not like commercial databases. As a result there was a big mismatch, incredible amounts of problems. A lot of the production came out of ways they hadn't planned on at all. As a result, with the LTER, each site has had a lot more flexibility that they can do good science and you don't need to be on the top saying "thou shalt do this important science". Instead you can let a lot of the good ideas come from the bottom and work up, not sort of top-down." (IM) 95.

… resulting in an understanding of the importance of preserving site control of the data:

"then the November '96 Information Management meeting where we actually came up with the idea of this harvest notion, harvesting to a central site." (IM) 96.

"you have a harvesting system, goes out to the sites that store the data and know what it is and they use it and they know what's wrong and what's right. That was a really creative solution…" (PI) 97.

Most recently, information managers have engaged in research in ecoinformatics, i.e. developing generic tools for ecology through innovative technology development projects where the challenge of simultaneously accounting for the sites' heterogeneities is more pronounced that ever (cf. Hanseth and Monteiro 1996).

"… anything after the process of acquiring data that has to do with managing data, analyzing data, modeling data, mining data, um, doing theoretical sorts of simulations, stuff like that…so sometimes we call it ecoinformatics and sometimes we call it environmental informatics and sometimes we call it informatics and sometimes we call it bio-informatics and sometimes we call it information technology. But none of those terms are really significant, they are all basically the application of the technologies to facilitate the discipline and there are lots of ways, lots of areas in which that can happen." (IM) 98.

"You know I initially was thinking that the plan we should pursue is to just focus on a subset of sites that were maybe the most prepared. Let them create EML and then find out where the problems are and sort of get that done. But then you know they would be able to then be showing off some of these tools that allow you to exchange data between sites. And then it is a much easier sell to some of the other sites. But the truth be told nobody wanted to be left out. I mean it was, everybody wanted, you know, at the information managers meeting it was like, we said well which sites you know might be interested in participating in you know the pilot effort. And every hand went up." (IM) 99.

"... this separately funded biological databases project, in a portion of that was devoted toward helping design EML. That project was all about how, using metadata to develop information systems that can react to information that has been encoded and then decide what to do in terms of getting data and making data available. To avoid writing custom applications for each individual dataset." (IM) 100.

2.3.3 Sharing and learning within a community of practice

A Community of Practice (CoP) is a special type of informal community that emerges from a desire to work more effectively or to understand work more deeply among members of a particular specialty or work group. At the simplest, CoPs are small groups of people who have worked together over a period of time and through extensive communication have developed a common sense of purpose and a desire to share work-related knowledge and experience (Lave & Wenger 1991, Wenger, 1998).

"I have plenty to do here managing [site] LTER information management. And at the same time I do feel the responsibility to be part of the community that is, you know my brothers and sisters LTER sites, and LTER information management." (IM) 101.

LTER information managers share overarching aims, interests and motivations though technologies and strategies are locally different at sites.

"There is this curious dedication among LTER information managers to doing what we do, into getting data online, into looking at new technologies. Just meeting the objectives set to us by the NSF. These people are not doing it because they are paid well, it's more a matter of really believing in what we do. That's what makes the group special." (IM) 102.

"We made a greater impact as a group, information management. Even though it's kind of a loosely connected network, there has been a lot of intersite work. It isn't a really cohesive network where each site is run the same way. Every site is very independent and very different. The network is that we are all funded by the same group and have figured out a lot of ways to do the cross-site science. But of the groups within the LTER, the information mangers are the ones that really have added a network framework. We are the ones that get along better than the scientists do, there are big egos there. We make things happen. We can do a ClimDB even with effort and struggle but we were able to do it. It would never have happened with a group of scientists; they would not have agreed on it. So I think we are an incredible asset to the whole LTER program, and I think they have made smart moves in funding our meetings, and giving us funding for IMExec and the phone calls..… It is the people, the ones attending this meeting, who have been really dedicated to this whole mission and its really made the network a huge success I think overall. [Participation in this group] has been one of the rewarding parts of my job...enjoyable and rewarding." (IM) 103.

"The information managers group, I just find it's sort of a synergy and interests and in terms of groups that have sort of productive and interesting meetings. The IM meetings really do stand out." (IM) 104.

"I think the information managers are in a sense a better group at networking, sharing ideas, ah, maybe that comes from the fact that they are a little more disciplinary focused because if you tried to do that in the LTER network because of its diversity, the odds that your particular message would appeal to, you know, 99% of the researchers is not to high, actually. Ah, you know if you were appealing to a quarter or one third of the people you are probably doing really well. Whereas in information management, these issues actually probably appeal to almost everybody, ah, at least the people who are with it or the sites that are with it ... they are looking for general solutions." (PI) 105.

"I think in terms of activity at the network level, there is no question the information managers have got their, they have got their stuff together. I mean they meet regularly, they have group projects, they inform each other in ways that ah, even within a discipline of the sciences it just isn't happening. The scientists just aren't motivated the same way. It may come from the fact, you know, not to explain it away, I'll bet you most of the information management folks are really mostly LTER, I mean that's their focus. They have been hired to do that or are assigned to do or whatever. The scientists are actually trading off multiple things; they're not only teaching, and advising of students, but they also, I think a good scientist would divide well maybe I'm wrong about this, but I think, by mixing up approaches so some, not all, my science is LTER science. I think that is true for a lot of them, so you're pulled in several different directions. If it was all LTER science, well you would have one less conflicting thing." (PI) 106.

"they [information managers] are a group that really has, I think, is a network group ... and one of our strongest network groups is the information management group. They really have been, often more than the scientists, which is kind of surprising I think…. you would think that in some ways that the scientists would be leading all these things. But they are not, in some ways they are being led and forced by NSF, information managers, by others who, perhaps they are not as engaged in the science but they can see the need more clearly. Maybe that is what it is, or maybe they have, maybe it resonates more with what they are trying to do. I mean one of the challenges of the IM is to make the network communications system actually work. So they are looking for examples, they are looking for successes. We have that challenge in the science front, but it's been kind of hard to get scientists to lead it. Some of that may be related to funding. I mean for information managers this is a big challenge. NSF and the network office have funded them to get together, ah, to go through some of these projects. So maybe that is it. It's probably a little harder for the individual scientist or group to get money for that because, basically you have to write a proposal and it has to compete with many other proposals. And that doesn't, they don't always fare well." (PI) 107.

Information managers see the IM Committee (IMC) as a community for informal sharing of ideas, mentoring and learning together. Information managers learn from each other's experiences with technologies through dialogue and joint projects as well as through attention to informing and training new participants.

"LTER information management… I have learned a lot from the broader group." (IM) 108.

"In information management I have learned a tremendous amount from this group. Right from the start it has been very rewarding for me to come to these meetings and learn from other information managers." (IM) 109.

"… a lot of it is our information manager group within LTER meeting on an annual basis and upping the individual's strengths of all the people. I mean I have learned a ton from those annual meetings." (IM) 110.

"…and it's all like being mentored really by the overall group. So I see we are training folks" (IM) 111.

"It is a good group of people to talk to and interact with. … And it is good to see what other sites, how other sites are doing things, as either a contrast or as an idea to improve. Because all the sites are very different in some ways: either in having just one person doing everything or whether they have relational database where they're trying to get everything in. Or whether they are just, what

> level, yea, it is just interesting to see how it goes." (IM) 112.

> "Because actually coming to this meeting not only, they were not only the highlight of the year, as in professionally, but it is in these meetings that I find somebody that understands my problems. That I find the support, technical support that I need…I do some system administration, I have worked with servers, and I have administered servers. ... you can find people that manage data but finding a person that does everything and on top of that is so aware of the need of documentation of the data and preserving the data for future generations, nowhere in the university ... You know I can find administrators, database managers but nobody does what I do with the vision that my program gives to me. So my heart is really with the program, the LTER program." (IM) 113.

> "They [colleagues, information managers] are always there for you. And even if they cannot respond to you at that moment, they will eventually. I am only but human, and I need that kind of support." (IM) 114.

The CoP provides continuity, shared awareness of long-term planning, a venue for building long-term relationships, and a safe place that is both a sounding board for ideas and an arena where IM voices can develop:

> "It is also true that we have sort of a core of people who have been doing this for awhile. So going to a meeting and seeing familiar faces and not having to rebuild the whole thing, you get a feeling of trust." (IM) 115.

> "I think it is learning and I think it is also, it is a place where people can let their hair down…be themselves, be natural. And it is safe to say things that demonstrate, I think failure is too strong, but where people have not been as successful or disappointments. And as soon as you are able to do that in a group, there is a bonding that occurs. And so it is not just a classroom setting… but it is also, 'YEA, I have that same problem'…" (IM) 116.

> "So the information managers network, clearly what makes it work, is the glue of these people who, maybe it's because at their site they are just the data manager, and when they get together with this group their opinions are considered a little bit more important than they might be back home. I think that might be the case with some people. For others it's a longing to have their job defined as something more than data management." (IM) 117.

> "I find a lot of stimulation and that's why I participate in this network. So the fact that we have some money available to do these meetings, on frequent basis, and the fact that we take advantage of it, I think is what makes a network a network as far as the information management system goes." (IM) 118.

Participation in the LTER IM Committee gives information managers a special point of view to gain an understanding of the LTER network level activities that PI's may even lack.

> "…with so many people in the network, well over a thousand, it is very difficult for all the people to interact at this level here [CC meeting]. We have about 35, 40, 45 people here; that is a very tiny fraction of all the LTER scientists. So most of us that come to these things on a regular basis, well we know everybody, we have come into a collaboration on common and group decisions. … But that means that the rank and file in the project oftentimes have a somewhat distant relationship with the national network level and that is one of the reasons why the All Scientists Meetings that we have about every three years or so are so important, it is the opportunity for people from the sites to get involved directly with all the other people from all the other sites" (PI) 119.

> "I have a gut feeling the whole network doesn't work as much like a network as the information manager network does." (IM) 120.

2.3.4 Sustaining multiple memberships

Information managers are located in a variety of institutions and participate in a number of communities. They construct networks to support their work,

> "I communicate daily with several members of the IM community, people in informatics

community, government agencies, international people." (IM) 121.

"I am an LTER site information manager and coordinate across two university campus computational departments. I work with the oceanographic and earth science communities in addition to the ecological and information management communities. Deciding which professional organizations to stay current with is an ongoing challenge." (IM) 122.

"I represent the … site in the LTER information management, commitment to annual meetings, larger IM group, IMExec…resources within our work environment: University Forest Science Dept, Forest Service, Fish and Wild Life people, USGS group." (IM) 123.

"I do also some liaison types of things. Recently ... I went to a metadata workshop to work on network level activities, all scientists meeting at our site (making sure there was enough food and hotels rooms etc.), [another site] advisory board, …workshop in wireless communications in D.C. … LTER schoolyard activities and technical issues, informal training of staff. … We had a week-long ILTER workshop in Hungary ... I've participated in several training workshops in the ILTER, also with OBFS." (IM) 124.

Information managers traverse between their sites and the LTER IM Committee, they have a foot in both communities. These memberships give them the advantage of multiple viewpoints and more extensive experience-based understanding of networking:

"The great thing about being an information manager is that you get to play both sides of it. As site person: oh man, look what they are making us to do now. And yet in your heart you know it is so good. And when you are with the network you can say: I'd love to do all, but I am anchored to this here and now, to the science. It is the strength of that position." (IM) 125.

2.4 Juggling multiple tasks within multiple time scales

From management to services to design, a multiplicity of endeavors are part of the everyday for the information manager. The multiple roles and memberships unfold in different ways but as integral elements of the information management trajectory, they are blended into a site's information management strategy. In the technological realm, articulation efforts are required to explore the major new issues faced by information managers today. Strauss and others show for the medical realm that articulation work "must be done to assure that the staff's collective efforts add up to more than discrete and conflicting bits of accomplished work" (Strauss et al., 1985; Maines, 1991).

Not everything can be addressed at once so there is an attempt to ensure that nothing is lost if an issue must be addressed at a subsequent time:

"And, you know… I sometimes get discouraged about it in that you know, I feel like I have, you know, I can only juggle like three balls at a time and I have 20 balls. You know, so I've got 17 sitting over there, and here and there, and I can only work on them 3 at a time, and so I feel like… oh, one of these days I'm gonna get caught that I haven't done all these things. You know, but, then the other part of me says I couldn't have done all these things. (Laughter) ... I just try to make it that, you know, nothing is unrecoverable. So that's my goal. My goal is to make sure that nothing gets like really lost." (RA) 126.

Information managers balance these tasks within multiple temporalities in their work. As the sections above have illustrated the tempo of information managers' everyday practice varies.

The complexities of temporalities relating to data are evident as data-related activities

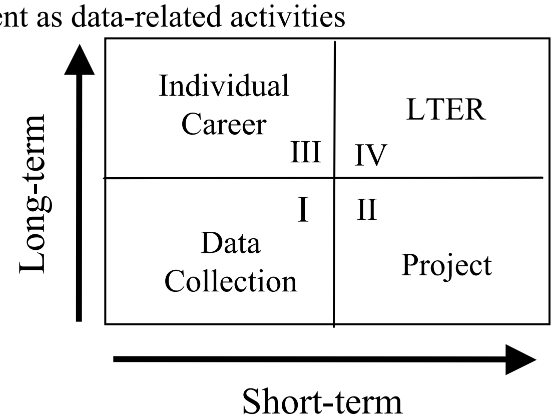


Figure 2. Multiple timeframes displayed as two dimensions (from Karasti and Baker, 2004)

appear interwoven at multiple time scales, crossing boundaries from the short-term to the long-term arenas. Figure 2 portrays the notion of interrelated short-term and long-term timeframes having to be addressed simultaneously since different activities occupy different quadrants. It represents data taking as occurring in the immediate present (Quadrant I) while digital data handling takes more time, extending into the short-term or project Quadrant (II) timeframe. Data analysis and information integration take place over longer career timeframes (Quadrant III) while decadal environmental data concerns reside squarely in the zone of long-term (Quadrant IV) labeled LTER.

There are predictable elements for an environmental data manager that include the immediate-term issues of seasonal and annual cycles of data collection, entry and preservation; the near-term issues of data use and publication (resulting from the two year data policy stimulus for scientists to submit their data and metadata); and the long-term issues of development and synthesis. Just as scientific design is present at all temporal scales, so too is technological design: there's an immediacy to field equipment operations or forms while there's a deeper vision at work for data integration and interoperability. Operating within overlapping timeframes, requires an agility to test approaches and products; a flexibility to adopt tools and procedures; an insight to balance maintainability and functionality of systems. Issues may shift between timeframes, i.e. security once seen as a problem for the future has become a dominant aspect of all contemporary computational environments. Yet the importance of the data is a constant and hence the attention of the LTER information manager to the long-term care of the data itself.

Juggling multiple time scales is an inherent part of an information managers everyday work practice. The skillful attending to various tasks requires ongoing triage and prioritization while immersed in the work as described below by a senior information manager:

> "One of those things that is true, that there is always more work to be done that can possibly be done with the resources that are available to you, and the people you can afford to hire. And especially when I first started my job I found that very difficult because there would always be some things that I thought that needed to be done that I could never get to because I kept having to do triage every day, and decide what was the most important thing to focus on, and set priorities. Eventually I came to some kind of peace with that because I felt that was part of my job, to prioritize and decide what was going to get attention and was not going to get attention and occasionally to acquire more resources. There was something that felt like it was really crucial and just had to do it. …"
> (IM) 127.

Partly it is a question of learning to deal with, plan for, and anticipate the heterogeneous aspects that require attention and resources at different time intervals for varying periods of time. For example, one information manager explains having been able to bring together the updating of web pages with preparing for site reviews and proposals (see quote #195), another one tells about having found uninterrupted time for writing publications while away from the office and site (see quote #57). However, when and how to migrate elements of an information environment remains an uncharted decision process, and often impossible to link with some established or fixed time scales. As there seem to be no existing tools and techniques that could help in bringing together the multiple time scales and various demands, developing this juggling skill is part of information managers' tacit knowledge. And still, at particular times even an experienced information manager is under a lot of pressure when activities relating to multiple time scales collide:

> "It is just a lot of challenges; I am working with four people, I have those people, you have the PI's, you have contractors, and then you have the general public. You are just getting hit from all sides. It's frustrating. I feel like I'm getting spread too thin. For one thing, I definitely can't concentrate on what I want to do! … At the same time we are transitioning our whole design so we are really caught in a big transition right now which adds to my frustration and freaking out, having so much to do. Then it stabilizes again." (IM) 128.

3 Relations and tensions in IM work

> "It is important to recognize that technology is a tool, and shouldn't be used as the end to itself. To recognize the benefit it can provide without simply using it because its new technology...does it really protect the information that you are steward of? And I think that is potentially a danger in having simply technocrats as information managers, without proper coordination and interaction with a science base. If you leave it all up to a technocrat you may get something that is non-functional in the end. "(IM) 129.

This chapter discusses the relations and tensions in information managers' work that the above quote illustrates. Information managers provide support for science, data and technology as part and parcel of their everyday work activities. Figure 3 illustrates an information management model designed to make visible the multiple facets of information management. These elements are strongly interdependent, but they are also conflicting in many ways. As a result, information managers may benefit from awareness and participation in the articulation work of balancing tensions, recognizing relations (Strauss et al. 1985).

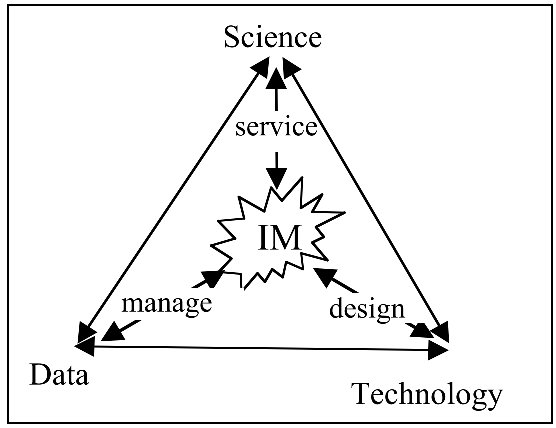


Figure 3. Relationships between the role of information management and science, data, and technology (from Karasti and Baker, 2004).

3.1 Providing support for science

LTER information management focuses on providing support for site science, that is, for a research team united by holding in common an ecosystem and a field site. In the words of an LTER information manager:

> "One of the things that I think is really important is that information management be driven by the research and that the information managers continue to come back to assessing whatever projects they want to develop in terms of whether it's really going to support the research at the site." (IM) 130.

As part of a community concerned with long-term issues, LTER scientists are engaged in ongoing discussions about data management and have developed expectations with respect to data issues. Senior scientists have a number of traditional insights that are impacted by change. For instance, data work is viewed as a single task of data maintenance to support research although the role has expanded to include management and design (Figure 2). Additional information management is viewed as successful when it is transparent although more participation is required to keep site technologies functional and network activity aligned. Data structure is considered with respect to its usefulness to ongoing science work yet is also an active field of research.

“[Data management]...to take all the messy data and get clean and make it available to them. (laughter). But in a sense I think that there is that naiveté, what’s all this money going for and what do you get for it ... And they simply don’t appreciate the time and the energy and the effort required just to do the nuts and bolts maintenance. Never mind any grandiose new stuff … I think the scientists come in a range of flavors, there is one flavor of scientist who would like all the information management to be totally transparent, the less they have to worry about it and the more that they can get from it the happier they will be. And there are some who take an interest in it and are party to proposals and efforts to both get money into the game and to make advancements in it per se. There is a full range of expectations.” (PI) 131.

“Requirements for data management were at first considered, there were people who thought it was important and people who thought it was completely unimportant, and there are still some people in either camp and I think we still have, well I told you before I think we have a conflict now between the people who think data management of a certain kind is particularly poor, and many of the rank and file scientists who have been around a long time and who are interested in, who have all this information collected over the years and are interested in being able to use that particular information in an effective way. And there is still a big tension between those who want to design the network information system and those who just want to use the information effectively being collected. … but I think we are on the verge, I think, of there may be a change. There is a growing interest in good information management that is coming from the ground-up at the sites that results from the need to manage the information that is already available. And I think that those needs and those desires are very different from the more top down approaches. … when we start talking about priorities or something, there is a large camp who feel that the data management is an absolute priority, it can’t be compared with the choices one makes among research objectives or in terms of research versus education, social sciences, and uh, there is also, I mean the arguments we had over network information system tended to, we get to the point where defenders of network information systems say this is the system we have to have and others say, well, do we have to have so much of our resources devoted to that system now. You know, can’t we take it a little slower. Can’t we do it a little different way. And there is a communications problem in that defenders of data management often sort of think of their work as absolutely necessary, a place where a compromise can’t be justified. We may be on the verge of changing that largely because the need for good data management at the site level are becoming obvious to the sites. So my hope is that eventually we will just internalize all these needs, and we won’t have to think about data management as something separate. It will just be part of the research.” (PI) 132.

“I think at the level of the individual sites, um, it’s hard for me to quite figure out. I think the scientists depend very heavily on the model that the scientists have of the information management is really different than the one that is reality. By that I mean in the old days when scientists collected the data, they put it on a piece of paper, they turned it over to the information management person, they entered it, they proofed it, they got rid of all the bad numbers, they produced the statistical analysis of the tables, and the scientists wrote the paper. How convenient. There is no money to do that now. The information management people have to kind of create the structure that will manage it, they will do some of the harder management, but the scientists have to enter the data, they have to look at the data, they have to tell you know what the units are. All the metadata stuff has to be done by the scientists. They are the ones who know what it is, and yet I see it at my site, I’m sure it’s at other sites, scientists operate under this old model, where you just turn over some crappy notes, and you know, paper and suddenly they do everything but write the paper, you know the final paper. It doesn’t happen that way and we have run into this conflict over and over again. And because of that they kind of, there is a conflict, well you’re not really doing what I think you should be doing. But those of us who ought to know, in the science end, know that model is like from the 70’s or something. It’s like make a central data bank thing, you know, that has been dead for decades. And ah, so there is an expectation” (PI) 133.

“the scientists, they get upset if something’s not there when they need it. But they’re not

> always thinking ahead whether they'll need it, whereas the information managers, part of their task is to anticipate what will be needed, and how to make it available quite rapidly. And the scientists, are kind of assuming that it's just going to happen" (PI) 134.

The IM views on supporting science develop as part of community discussions:

> "Probably something most sites have problems with, while the intentions are good, there are gaps in communication between researchers and information managers, keeping the information management up to speed on what's going on or including them in the research level." (IM 135.

> "You want the investigators to be able to put in their time doing the science, and spending the least amount of time possible doing reporting and all those types of things." (IM) 136.

In response to these expectations pertinent to particular scientists and sites, information managers see that:

> "You [information manager] have to … be willing to, to some extent accept a support role to the main scientific function of the LTER." (IM) 137.

> "at a certain level my job is to represent what the PI's want. I mean they are my customer. And so if we can act as guides to get them where they want to be …a very consultative process." (IM) 138.

The science support focus is also reflected in a senior information manager's view of more junior participants:

> "I have listened to a few newer information managers and they are kind of all set on setting up this particular system and its going to use this software and its going to this and its going to do that, and they are getting resistance from their PI's. And so that seems to me to signal that they haven't done enough communication, they haven't allowed their priorities to be influenced by what is really important to people doing their research. So I think you need to keep coming back to the primary reason that you are doing this information management is to increase the research productivity at the site, and so that influences your priorities all the time." (IM) 139.

3.1.1 Support strategies

Information managers resort to multiple locally relevant approaches in support of science. Three of these are described below.

Keeping IM transparent – service relation

A not uncommon assumption held by scientists is that information management should not cause extra work for them. If it does cause extra work, it slows the scientific work so IM is then less likely to be done at all:

> "it's just time is a premium. You can't afford an extra 5% time, let alone 30% time. Just not going to do it." (PI) 140.

At a number of research sites, information managers try to live up to this expectation by keeping IM as transparent as possible. The following excerpts explain how they actively hide some of the technological and theoretical complexities from the users (thus also making their own work invisible):

> "So I think you need then to return to them [the researchers] really positive things and make it as easy as possible and try to hide as much of the complexity of the [information] system to them as possible." (IM) 141.

> "I am going to provide them with a metadata that they need, in order to be able to do a better analysis. I am not even going to call it metadata because they might get scared." (IM) 142.

> "Because relational database required that you have very clear set theory. And if you don't you are going to have a problem. Of course you can learn about queries and stuff. But to design queries you need to have that. And that of course if you start talking to that, about that to an investigator, you will lose them. They don't care about that. That is why you see all these systems being developed. The web tools that have been developed so you make, you allow the investigators, or the users, to make queries without even knowing that they are using set theory." (IM) 143.

"but it's always a question do you spend time doing something new or do you spend time revamping behind the scenes that does not affect how the PI's see it because that is really the important part, the users. Are they able to get what they need out of it." (IM) 144.

Educating participants – reciprocal relation

Information managers recognize the impact on strategies brought by focus on long-term data collections and the importance of communicating abut the data. Since information management is embedded within the science community, there is a constant flow of information about the local science research so that the IM is immersed in and attuned to the developments and needs of science.

"the communication that goes on between information managers and the researchers, is a two way kind of thing. " (IM) 145.

There has to be that two-way street between science and the techie so that the service that has been provided serves the needs of science as well as providing the protecting cocoon and the ability to service that data to others outside the community." (IM) 146.

"Information Managers are always invited to PI meetings and encouraged to participate, encouraged to share developments, or how things are going with specific projects within information management. (IM) 147.

They describe the process of interaction as 'informing and educating', 'change agent', 'collaboration' and 'two way':

"It is a constant battle for the information managers to keep the PI's informed of the types of things we are doing. Some other people like the other information managers know more about our systems than some of the PI's do." (IM) 148.

"Information management … the invisible part in the sense that it is hardly ever spoken about. But information managers seem to engage in this at their sites, educating the PI's and being the change agent and those things." (IM) 149.

"In LTER we are changing the model. The scientists at LTER didn't train in LTER sites where there is a lot of collaboration, and information management was a big part of that. As grad students they were out there collecting their own data, analyzing their own data. When you're done, your data goes in the file cabinet. Fifty years later you retire and someone throws away your data. There is a lot of teaching and re-educating people." (IM) 150.

"Another aspect of the role of information management that I do as well, it's not something that is relegated to the office of handling data or metadata, it actually begins with the interaction with the researcher before they initiate a study to help them structure the information they are going to collect so that it can be readily integrated with an IM system as well as having it in a form that is readily useful by others. Establishing that contact with researchers before they begin their study. Using this notification research application, which also includes a section that they sign or acknowledge that about sharing the data and for providing metadata for the information that they collect and making them aware of expectations ahead of time so they don't have any surprises of what is expected of them or the information data that they collect." (IM) 151.

"So, it is important to think about how you can broaden the researchers and the graduate students view of information management. Usually once or twice a year I ask for some time at our monthly science meetings to present something related to information management. I just did a presentation a couple months ago on information management at the network level because I thought that a lot of people were familiar with our website and bunches, a lot of them have taken our training course for how to use the database browsing tool that we have, so they knew a lot about our local system, but I think a lot of them didn't know about NIS modules and about EML and the directions that we are headed, and about efforts to build infrastructure support for inter-site research." (IM) 152.

"It has required a great deal of education of the researchers, getting them to be aware and recognize the importance of documenting their research and the data they collect, and why it is worth their time to do that. For many of them it was a very new thing, they did not recognize the benefits.…" (IM) 153.

"One of the roles the information managers have assumed, over the last, well the years since I have been involved in LTER, is helping to educate the scientists about the role of information management and its importance in the LTER enterprise. One of the, I think many of us find ourselves coming to these annual meetings finding those feelings reinforced by common experiences and group, and being able to take that reinforcement experience back to the individual sites and working with the scientists. And I know at our site that has been a tremendous evolution of understanding about information management over the years." (IM) 154.

Pricing IM services - requiring recognition

Some information managers have emerging thoughts on pricing the services of information management:

"I have tried to educate PI's that we are a resource, just like a natural resource that, you know that you can only, we are not just free. And I think what happens, is because we are funded for so long they don't necessarily, the PI's don't necessarily equate the fact that if I have to do something or one of my employees has to do something, that there in fact is a cost associated with that ... every time we get a request for something, I always talk in terms of OK it will be this many man-hours for x personnel, and based on their salaries this is what it is going to cost …." (IM) 155.

3.2 Providing support for data

Ecological research typically deals with heterogeneous data and poses for information management the challenges of dealing with long-term maintainability and data diversity as we have described in section 2.1 (Managing Data). Information managers are motivated to provide support for data because they are aware:

"I just think that the questions they are asking of the dataset have changed over time. As [the dataset] becomes a longer resource, it becomes more and more valuable." (IM) 156.

"They are finding that the data are becoming more and more valuable as the years go by. And, so again even the, sort of the experimental scientists who work on the present moment, present day measurements, present moment measurements are getting this sense of the importance of having a long-term dataset. I mean that is the whole purpose of the LTER, but sometimes that doesn't reach into all fields necessarily. But it is starting to now…" (IM) 157.

3.2.1 Data stewardship

With managing the growing expectations and attending to the growing complexities of information interfaces and technology, there is a danger of neglecting the data. As science researcher needs are heard and new technology is supported, these concerns can compete with and alter long-term plans for data care. There is danger of compromising the data when immediate issues take precedence: PI's make their needs known and agencies proclaim technology requirements. 'Data care' is a longer-term, chronic task essential to maintenance of core datasets and recovery of legacy data.

Data stewardship is manifest both in the 'data care' practices that LTER information managers provide for gathering data and metadata to assure data quality and in their activities of designing databases and technologies to support long-term data maintenance (see 2.1.5).

Databases are more tangible entities in contrast with the largely invisible, sometimes quite mundane, day-to-day data processing work itself:

"After that you have to go through some quality control and double checking and it requires you're patient…dedication to that. And it is boring." (IM) 158.

"…besides providing an infrastructure …providing a structure. When the investigator comes to me at the beginning of the project, I help them organize the data, even before they start. I have a set process of us entering the data, having quality control, double checks on that data. Two people the one that enter and another … checking the data that has been entered. And given a first report of the data to the investigator. The investigator is supposed to double check that data, you know. Send it back to me. I incorporate the changes and send him or her a second set of reports. And he is supposed to go again through the data. The datasets …in better shape are those that go to a third process of incorporating more data and the format being manipulated. You know the structure, when you do querying or sometimes you have to extract certain columns from your dataset only or do summaries or sometimes you have to transpose everything and change the format and the structure of the data set because they like to see it from another point of view. That is where the real problems of the data come up. So going through that third report or manipulation of data really makes a better job. So in that sense I help them. But that is day-to-day dirty work ...” (IM) 159.

Attending to the day-to-day management of data flow often requires time consuming but unarticulated and therefore ‘invisible’ work for information managers:

“it is really hard to get people to take the time to read through the instructions, and to get their data into that format, and to actually participate. We can see the error messages ... like wrong dates and mins exceeding maxs and things that cause all of these error flags. Doesn’t anybody have data like in a decent shape? Some sites did a lot better than others, but it was really enlightening as to what state this ecological data really is. That there really is the need for more quality assurance programs at the sites. I think it’s just that we are all trying to do so much ... that people don’t have time to participate in these things.” (IM) 160.

“… they gave me the data. But they gave me the data in all, like in chunks, in piles and so it took me awhile.” (IM) 161.

“We really need to promote mechanisms for being able to get the information, get the data as well as getting the descriptive data, and the metadata is really the most difficult.” (IM) 162.

“We are actively working towards getting that knowledge about their data from them before we, you know, lose that data, they pass away or become ill, that type of thing.” (IM) 163.

Data care strategies are a part of everyday practices and long-term planning:

“The data has to go through a process too. … [It] is always better if you expose your data. And I always tell my data entry personnel: we are human beings, there is no way that we can escape from mistakes. We just have to be aware and minimize those mistakes, and you minimize those mistakes by trial and error. As you do the mistakes, you find out where you make the mistake and why. And that is the learning process and then you change the way you do things. And this is why sometimes you know people do certain things in certain ways because they have proof through time that is the best way to do it.” (IM) 164.

“the QA is a big issue, in terms of like curatorship ... one of the things you can do quickly is check for extremes in the data, like data ranges. So at our site we have in our metadata, and in the EML too, this is one of the things with the ecological metadata language, that they hope to do that could kind of be a shared routine. But you can describe like the min and max that would be common in a dataset. You can also describe if you have a value that is a coded value like you might have like a substrate code in one of those stream cross sections. You might say this is cobble, then it turns to bedrock, and then there is a boulder, or there is a log in the stream. So then those are coded, and then what you can do is then run the dataset against the master list of the valid codes, and it can tell you, you have a wrong code here, something is wrong with the data. You can also check the extremes, and it will tell you it is outside of the bounds. And then there is a number of other things like they can check for duplicates, where it shouldn’t be a duplicate and things that we can all run, the general test that just relies on reading it from the metadata and then comparing the data against that metadata that you know, and provide a report, and it works the same without having to reprogram. You don’t have to redo anything. So that is one of the tests we always run on every dataset that lets you know that the structure of the file is proper, things haven’t gotten out of alignment, and that you have all valid coded entries. And let's see what else. The other thing we added was you can add extra code, so we run this general program, but then it can tack on the specific code and actually run it as well that would provide specialized rules. And those specialized rules might then say, OK, it is going to look at the last time we measured this tree, and determine whether this tree you know was dead last time, and now is alive. Or had a ... diameter of say 20 centimeters before and now it is 10, that would be bad because these trees don't shrink. So there are all these rules, things that they know have to be true. And some of these we just discover ... well we should check this ... and so the more you know about the nature of the data, then you can program in all these other rules. And that gives you really thorough testing." (IM) 165.

Due to its contingent nature, information management work often has elements that are not part of a routine sequence of events whether gathering data or designing databases and technologies for data maintenance.

"I am going to help them standardize those key data points. So at the end, when they get to the point to analyzing the data, then they don't have to be translating, you know the name they gave to the plots and the format of the data. Because I know that if I allow them to do that, we are going to have as many coding methods as key investigators" (IM) 166.

"We had to create new numbers because we went over 10,000 and little aluminum tags that you put on samples out in the woods go from 1-9999. So we actually had to CREATE a new row of numbers because we went over 10,000." (RA) 167.

"We had to add a comment field because as time went on samples would come in ripped or torn or full of bugs or whatever and we had to find a way to make that known. Then I had to add in a flag field for the outliers. I had to find a way to mark them because we don't want to delete any data, we just want to mark it." (RA) 168.

"They had a different version of the web server. They had a different version of PERL ... They had a different version of the database software itself. And every time you change one of those pieces, although everyone says it is interoperable, it is not." (IM) 169.

"The new metadata standard asked us to have a method associated by attribute. It's one of those things that the expansion of the metadata system is causing us to connect. This is going to be very difficult to maintain ... because it's very comprehensive. I like it but I am worried, hard to keep it up to date." (IM) 170.

A site information manager traditionally has both an intimate, local understanding of the data together with a deep knowledge and concern in the handling of data.

"The thing is that I have been there for a long time so have developed all those systems that dealt with the climate data and particularly the stream flow information, the stream chemistry." (IM) 171.

"in my case, I have kind of been, we have had these stream flow data, and our climate data, and like some of the survey data that we do of stream profiles, looking at changes in the wood and the streams, and changes in just the substrate, like boulders might get moved downstream and it changes the pool and ripple, anyway there is a lot of that survey ... it has been going so long all the PI's that are here, none of them are the ones that originally started it. So you have all these PI's ... that kind of get assigned, you say well, you might have somebody come in …[and want] to use the dataset so they are assigned sort of as the PI for the dataset since it has been going on so long. There is really no [original] PI there. But they don't understand how it is processed or anything so in our case I end up doing all of the maintenance, all of the metadata, everything. I mean that dataset wouldn't be a nice long term dataset if it wasn't for the data manager." (IM) 172.

Information managers are well positioned to design databases so that their relations to other databases, schemas and technologies are taken into account.

"He (a PI at a site) would be better off if he had us review his design, but he always thinks he's perfectly capable, knows perfectly well what he's doing. Actually he does, WAY better than just about everybody else ... but a lot of times his stuff also needs to be reformatted when it comes into the lab. But he, at least, is thinking of designing forms from the field that are nice and readable and can be key-entered by key punch people. And a lot of other folks, they don't have good forms. They may enter it themselves into a spreadsheet that's disorganized and it's not designed in any kind of manner for a database ... and that's really one of the reasons [name of an experiment] I think has been very successful, is because going in [he] recognized that the data management effort was going to be very significant. I think he still underestimated it, but he did recognize that it was going to be a massive effort … I think things could have been a little more timely, they, the databases that were built could have been built more relationally structured from the very beginning. We had to kind of go back and build a relational structure ... Most of the mistakes were fairly minor, really. And they were ALL, they were all able to be corrected. I think it just could have been a little better designed from the beginning." (IM) 173.

As groups get larger and expectations grow, there is danger of loosing intimacy with the data and the technology.

"And as our project has grown, I have moved into a much more administrative kind of role ... And in some ways that can lead to me getting more out of touch with, actually being able to do the technology." (IM) 174.

"You try to meet the new expectations. Of course it is exciting to go work on a new design. Cool community to collaborate and work within ... But it is moving up through the matrix [... and] away from the square that already demanded attention (the data). Maybe it is that you forget that the integrity and quality of the data, does still depend on – not the instrument (only) but – the individual and partnership with the instrument." (IM) 175.

"Some sites avoid increasing distance between information management and data by having the data manager work part time as a technician.." (IM) 176.

Keeping track of the data

An information manager's role includes responsibility for not losing track of the data or the information.

"We don't want to lose [data], lose the opportunity of getting it from them, so of course, you take it and you work with them to get it. But, you know, we're overloaded and they have different conceptions of how fast things will be up and how nice it will be. We'll eventually get them up." (IM) 177.

"You can't do it all. So you just have to make sure you don't lose anything. …And ah, you know, I can't do it all. And so I try to do the thing that's most critical right then. And a lot of things end up on my desk. And they don't get done, but they are not lost. (Laughter)" (RA) 178.

Extended temporal horizon of 'data care'

Long-term data concerns extend the temporal horizon: to the future as well as to the past. Information managers address the varied concerns in their everyday 'data care' work by:

- working with past data sets:

"I was trying to document a lot of historic stuff with pen and paper and just asked the PI questions… he was coming on with Alzheimer's and I knew that he was going to retire … and I had a series of interviews with him and I got INCREDIBLE docu, I mean, I got all the documentation for these early corporate [data], like stream chemistry and things, all from just doing interviews with him." (IM) 179.

"So historic data is an important aspect of the project and dealing with all the problems historic data is." (IM) 180.

- taking care of the current, ongoing data capture:

"getting their [scientists'] data into our system from the very beginning to, whether it is to help them with data entry forms, setting up data entry programs, all the way from you know QA/QC programs to getting it archived into our system and accessible on the internet" (IM) 181.

- designing data infrastructures for the future:

"as we envision it also that we'll also be adding the EML [Ecological Metadata Language] … And sort of often go back and forth between whether we want to do that from the ASCII files or the database. … but at any rate we'll somehow make EML available dynamically on the Internet to the group at large, to support EML in that effort for having a standard exchange format for metadata. That has really been my focus for the last year or so." (IM) 182.

Quantity - quality data tensions

Limited resources traditionally make evident in the field sciences the tension between quantity and quality of data. One end of the spectrum is to collect lareg amongs of data and the other end is to limit collection to fewer, well-tended valuable datasets.

> "We have data coming in all the time. We've an incredible amount of climate data that's collected. We have 4 benchmark stations...they collect data every 15 seconds, on at least 10 different climate variables, and a lot of other satellite stations just collecting temperature and precipitation. I think it is like 7 million data points per year, or even more. Just an incredible amount of data. There really needs to be QA/QC before you can really put it up." (IM) 183.

> "I think the other things is this balance of, between you know, sticking with it and not adding things and adding things. So the, to some degree the change is limited by resources. If you have more resources you could change more rapidly and could add things, but that, that could dilute efforts and thinking so, it's not necessarily all positive to just ever build the empire larger." (PI) 184.

> "I think in the beginning the PI's just wanted to go out and get the data. Got back to them, no time to do analysis. Enthusiasm, can't help it. When you go there [the field], you want to get all you can. Now they are backing down." (RA) 185.

3.3 Providing support for technological infrastructure

As technologies are developed at increasing rates, staying technologically informed is an important aspect of an information managers' work:

> "need for people to remain current in technology" (IM) 186.

> "Technology keeps changing, [our] original tape library and mainframe system, it was really kludgy but cool at the time. It is a constant battle to keep up with things". (IM) 187.

The impetus to stay technologically current is counter-balanced by a long-term perspective advocating a conservative approach in site information management systems. Long-term data management is, almost by definition, moderate in its approach to technology procurement.

> "When I started working with the LTER folks, I thought here is this group that has you know 15 or 20 years of experience doing large scale integrative research. They must have just amazing things going. And so when I went to that meeting ... I was just really shocked to see the kind of pitiful state of the structure of this network. It was only later that I began to understand that there were a lot of organizational issues that were … directly impeding their ability to, or even to desire to be a network. … It just seemed like when I got to LTER, they gave talks about like, um, mailing lists, or um, how important TCP/IP was, and these were things that were like, you know I sort of felt like those were 8 years ago. And, I don't know, it always seemed like in that first year that I was getting to know them they were always 4 or 5 years back. And just catching up. … like I saw a 10-year plan for LTER that talked about developing this network information system… And it was interesting to see just how long they were giving themselves to move across this ten-year plan. There were little baby steps to get things like email servers, and integrated stuff up, but it, I sort of felt somebody could put together and you know at least a first pass system in 3 to 6 months if you had the sort of network will to have it implemented across the sites…" (IM) 188.

LTER scientists who design long-term experiments also plan cautiously:

> "…and the key is doing something in a pretty fundamental level, in a way that is not too subject to changes and technology or perspective or anything. …" (PI) 189.

The persistence of technological change and limited resources prompts cautious thinking and careful balancing of options. The concern is first for securing safe and stable infrastructure rather than implementing the latest high technology innovations. Judicious decisions about technology procurement are influenced by the features of high reliability, easy maintainability, and low risk for long-term data management and science support. An information manager's foremost concern in aligning developing technologies with existing infrastructures and practices is to minimize disturbance of ongoing data archive and use followed by interests in optimizing long-term data use through updated technology.

> "that experience we have had with several of our things… that the issue isn't how you do it, it's how do you maintain it and how do you make it so that it is easily maintainable." (IM) 190.

> "And it is going to take just as long the next time unless you create the basis and infrastructure that you can build from and move forward." (IM) 191.

> "there are certain key infrastructure things that you have to do to get the job done. In fact, by not doing them, it may not only influence what you do now, but it could influence strongly what you do in the future." (PI) 192.

Almost ironically, information managers, though conservative in their technology acquisition approaches, are also the technology proponents, advocates and designers both at their sites and within the larger network. On one hand, there is the concern for having in place a data-safe, functional system. On the other hand, incorporation of new capabilities to enhance data capture, access, preservation, and analysis always holds the potential for further facilitation of science.

In addition to balancing the tension between the speed of technological change and the work of 'data care', an information manager is required:

> "to do long range planning when new technologies can be placed in, look for the windows of opportunity for proposals for major upgrades for technological infrastructure". (IM) 193.

The evaluation process that places research sites under scrutiny every three years sets a timeframe for some technological updates:

> "We manage to update it [web pages] every three years, for review and proposal. We are on this cycle, and we end up putting a lot of energy into updating." (IM) 194.

However, transitions of a larger magnitude occur less often:

> "Every so often things need to migrate, the technology changes so much." (IM) 195.

> "having the investment in [current technology], it is not so bad yet that I would want to go and rewrite all my interfaces." (IM) 196.

These ongoing and judicious technology procurement and implementation processes produce "a kind of archaeological layering of artifacts acquired, in bits and pieces, over time" (Suchman et al,1999).

Information managers adapt and align technology, data management and information flow at their site. With the procurement of new technologies they have to pay special attention to site practices, with existing technologies and with available funding possibilities. Furthermore, they are required to keep up with and understand developing standards and methods. In these balancing activities they need to be able to understand the local research because – as already stated - providing support for site science is the initial, immediate, perennial responsibility of information management:

> "it isn't a conflict between infrastructure and research, but in fact you enable better research by having infrastructure." (IM) 197.

In a profession with tasks that require constant prioritizing, an information manager recognizes there is a danger of neglecting the infrastructure:

> "There is some tension between things that are pressing versus kind of maintaining the infrastructure sort of things. So there might be something like the … paper, the presentation that is coming up and you know its got a deadline, and you get to your office, and there'll be e-mail and people want data, then somebody emails you and asks you if you if you'll review a paper, and you know so there are all these things that kind of are somewhat time critical, and usually a lot of those, and yet there are really some mundane level stuff of just even at the level of house keeping, cleaning your files and backing up your computers and all that kind of stuff that has to go on too. And since there is never enough resources to do it all, you kind of always having to try to strike a balance between the things that are kind of time critical and the infrastructure that doesn't have to be done today but you can't always blow it off." (IM) 198.

3.3.1 Support approaches

Information managers differ in how they deal with managing data on local site issues (Baker 1996, Benson 1996, Briggs and Su 1994, Ingersoll et al. 1997, Porter 1996, Spycher et al. 1996, Stafford et al., 2002, Veen et al. 1994) and at network levels (Baker et al., 2002; Henshaw et al., 2002; McCartney and Jones, 2002; Melendez-Colom and Baker, 2002; Porter and Ramsey, 2002; Sheldon et al., 2002; Smith et al., 2002; Vande Castle et al., 2002). The following examples describe some of this continuum and also illustrate site-based considerations for the chosen approaches.

Keeping it simple

A number of sites argue for "keeping it simple".

> "And our approach on the IT infrastructure is then to keep things as simple as we can and still provide the services that we can." (IM) 199.

> "these days as you see me choosing the low tech path as the one least likely to fail … there was a time ... that I would set up with this sort of situation and automate it. Now I am cutting and pasting as the saving mechanism did not work the first time. I know the finite number of files I have to copy and paste, know that it will be done in a half an hour, and if I try to figure it out ... to somehow automate it, it would take much more time. The chances of it being a similar problem next year would be very low because of the amount of change in terms of in the ways they are saving it, what they are using to save and send it. Computers would have changed or what is happening at this end would have changed or my understanding, would have changed of what I just did, and I would not know why. And so a lot of the automations proved not useful in the long run. I find my strategies are different now. I tend to do things a lot more manually now. … There are so many ways technology allows you to take these days, that it is matter of which ones you happen to have in your tool bag and can make work. It is not like it's the best way to do it, which is a big change, as I was telling you about automating, I used to look for the best ways. Now it even cannot be on the list. " (IM) 200.

> "then we just archive it in both Excel file on it, on our in-house database and then as a text ASCII delimited file with the associated documentation. And then that gets put on our web site. And our web site, it gets processed by just data file, documentation file gets processed by a script into a web page and then that points to the data file. So very simple and straightforward. … So in that way, it has sort of been our philosophy to keep things as simple as possible, plus that the person responsible for the data files, at least familiar with the area that they are working on so they can give some idea of what the quality of the data is." (IM) 201.

> "one of my favorite quotes regarding metadata, which is one from Antoine de Saint-Exupéry …*"perfection is achieved not when nothing more can be added, but when nothing more can be taken away"*. Because with metadata there is a tendency to sort of, you know, keep adding on features until all of a sudden the features outweigh the, or the cost

of the features outweigh the benefit of the thing. But at this point I am not, like I say almost, there is practically no metadata standard that you can throw at us that we are incapable of doing. And if in fact we're providing more complex information where less complex information would work, I can always re-lump it again. The danger is you limit participation. That is the bad part. If it is too complicated there are just fewer people willing to do it. But the LTER community has the resources to make it happen regardless, if we will spend the resources that way." (IM) 202.

Automating systems and experimenting with technologies

Some other sites count on automating systems and experimenting with technologies.

> "The philosophy for information management at our site is that the website is sort of the filing cabinet for the site. … people are widely distributed. … so my philosophy with information management is to make it so that as much can be done by the investigators and students as much as possible, so that I spend more time setting up systems that do things rather than actually maintaining them. Part of the job is learning new technologies, to try them out, so when something needs to be done, I know a good way to do it. I am still playing a lot of catch up with a lot of the XML, and some of those types of technologies. I haven't had a chance to install all the tools and play with them a bit. What I find is most effect is to get a basic set-up and do something that is very small, very simple and then improve it so it gets a little bigger and then a little bit bigger. Then [when something needs to done] it isn't a question of sitting down and writing something new, it's always just a question of taking something old and making it better." (IM) 203.

> "You know, we tend to be very good at sort of leveraging existing products and technologies in order to do valuable things for the ecological community." (IM) 204.

Data accessibility and exploration

Some others have emphasized data accessibility and exploration.

> "we put a lot of effort into making data accessible. My goal has been to have it be easily accessible. When the project started in the early years before we had a lot of the pieces in place, a typical scenario would be that a researcher would want a data set, they would have to come to the information manager, and then a day later have the dataset that they wanted, and I think that a system like that inhibits the exploration of data, to be able to ask questions as they occur, a really exploratory relationship with the database. We at our site have put a lot of effort to the accessibility and making sure that people could access the data directly and could do it in ways that they could get just the data they wanted to get."(IM) 205.

> "Like we could have invested tons and tons of energy earlier on in our project in developing elaborate metadata. But it was not my perception that, that would produce as much effect on our researchers at our site as putting energy into access to the data and building a real powerful system for accessing data. So we chose to put our energy in that direction because I felt the researchers at our site were doing just fine without more elaborate metadata system in place." (IM) 206.

3.4 Balancing further tensions

On an everyday level, as jugglers of tasks (see quote #126), a type of mediation occurs that establishes a site's particular choices that balance some recognized tensions.

3.4.1 Short-term - long-term

Long-term science is concerned with the research need to collect and keep records of the same measurements over long periods of time. At the same time it is necessary to attend to the short-term concerns of innovative site research and publications that are assessed at three year intervals and critical to success in securing the next increment of six year funding. A senior scientist observes the tensions between short-term and long-term issues and the implications for information management in providing support for science:

"some of the tension came from the difference between people wanting to use the resources for short-term business as usual, process oriented studies, versus maintaining a long-term program with a legacy of a database." (PI) 207.

LTER participants are often heard discussing strategies for balancing short-term and long-term work.

Data:

"But I'd suppose that the people who are sort of, hard core in the middle of the LTER, the people with the long term perspective, it doesn't mean they don't do the other work, but it means they have an appreciation for and ... they don't disparage long term work. They feel it as an important aspect of what they do, they see this as an important part of their own effort, LTER network. People who put time into building the data legacy I think realize that they have to have some kind of sense of perspective in that perhaps 50 years from now, some bright graduate student can take this 50 years of data and learn something that might have otherwise not have been learned, that everybody has that sense that will certainly be valuable." (PI) 208.

"I mean just the notion that you would be collecting information over the long term means that you really have to manage information. If really that is the nature, to provide information over a long term, then you are going to have to provide management for that information." (IM) 209.

Infrastructure legacies:

"It is always ironic, that almost always the most technologically advanced site for information management within the LTER network is the newest site, the site that has just started. The reason for that tends to be that you can start fresh ... pushing the technology the furthest." (IM) 210.

Newer sites do not have institutional legacies, old technologies or large quantities of historic datasets. Alignment becomes more complicated as legacies grow especially as there is an increase in interdependencies between data and technologies, between organizational and institutional practices.

3.4.2 Site – network

Increasingly, an information manager for an LTER research site balances support for a site's ecological research and associated network activities. Providing support for the site science is the initial, immediate, perennial responsibility of information management (see quote #130). The site and network science with differing goals and scales constitute a dilemma, creating a tension prompting continued reflection, evaluation, adjustment, and negotiation.

"When LTER information management started, the focus was on the site science, on site information management. The information needed to be shared within the site. And even that was a revolutionary idea. At the same time people were in contact, can trust each other. That was a lot easier than opening the door to people from outside the site." (IM) 211.

"But when you look at the actual situation at each site, they are swamped with things to do and there really isn't enough resources for them to necessarily take all this time to participate. It is a big problem. It leads to the turning over of data managers at each site ..." (IM) 212.

"where should the information manager's time be going. Should it be only to support site activities or should some of it be going to support network activities or should some of it be going to research?…It is a balancing act; it always will be. It's my primary responsibility is to the home institution and that has to remain that way." (IM) 213.

3.4.3 Local – generic

Information management responds to both local and large-scale influences. The local customization of information management work evolves in tension with the development of common standards (cf. Hanseth and Monteiro 1996).

"to realize that even if a powerful system is there, simple and straightforward, spreading it to the community is a whole different issue." (IM) 214.

"There are differences that are small that you could improve with standardization but we pay a high price for standardization." (IM) 215.

"EML really excites us from the sense that we can get away from having to develop our own internal software applications to a large degree and actually leverage the ecological community at large to provide higher level tools for analysis, and modeling to our researchers as well as to make our data and metadata available for cross-site studies." (IM) 216.

"[PI's name] and I do have very different visions of information management and the role it plays. You know [PI] makes the, to my mind, reasonable demand that before he wants to sort of invest in information management, or wants to sort of see information management research, he wants direct evidence that it is benefiting the ecological research. And that is by no means an unheard of opinion out there. And I think, you know, it's a long-term ecological research, not long term informatics program. And you know we really do need to create products and do things that can have direct impact on the research. And I think that within, at the site level we do have that. And getting it so that it works at the network level." (IM) 217.

4 Information managers' work in change

The everyday practice of information management comprises a thickly interwoven texture of technological and social threads, requiring skills and knowledge in various areas. Furthermore, there is ongoing change and increase in expectations.

4.1 Varied tasks require diverse skills

"My work is fairly broad ranging." (IM) 218.

"It's just a bunch of different things." (IM) 219.

"I wear many hats." (IM) 220.

"I do a wide range of things." (IM) 221.

"So my day-to-day tasks are highly varied." (IM) 222.

"A site information manager must be a generalist." (IM) 223.

The above quotes vividly describe how the work of LTER information managers contains multifarious tasks and responsibilities including an understanding of ecological research as well as social, communication, and technical issues.

Science skills:

"Absolutely critical: you have to have enough understanding of the science and familiarity with the scientists and the type of things that will work for them or not work for them. Because there are some classic examples of building technological tour de force systems that scientists take one look at and say, 'we don't need that'." 224.

"I think I have been helped some over the years by being, working on the scientific side of things, so and many of the information managers have, and I think that it helps you a lot in understanding the perspective that the principal investigators have." (IM) 225.

"Certainly it helps to have an understanding of ecology and kind of a scientific enterprise. What is involved in doing research and getting funding and writing publications. All that is really critical to understanding how principal investigators operate and how the funding works for the LTER program. (IM) 226.

Social skills:

"main skill is people skills, but of course the professional skills are very important." (IM) 227.

"I think a critical part of it is the social aspect of being able to work well with a whole range of people. Some of whom are very easy to work with, some who are difficult to work with, and everything in between. You have to like people." (IM) 228.

"Interpersonal skills are very important. If you are going to convince the investigators ... interpersonal skills are a good thing to have." (IM) 229.

"Although work needs to be done, I mean you always have lists of priority of things that get dropped off if you don't have time to do.

Probably the biggest tensions are just interpersonal tensions between different personalities." (IM) 230.

Communication skills:

"Helps to be a good communicator. To be able to write and speak effectively. I think that is very useful." (IM) 231.

"Communication skills can also be used, not just interpersonal ... given that LTER is building something new, getting people to think about information management. I like to think every information manager is out there spreading the word, being able to do good presentations, being able to write papers, being able to communicate with the written word as well as spoken one." (IM) 232.

Technical skills:

"You need some ability with computer technical matters. Obviously want to be someone who is comfortable with computers and with numerical techniques." (IM) 233.

"Well it involves archiving the data, documenting the data. Writing programs, or data entry programs, web development, database development, database management. Writing proposals and grants. This has become a larger part of my job recently. Networking systems, PC support, software installation and configurations. Getting them to interrelate and that type of thing. Pretty much I do a little bit of everything." (IM) 234.

"In the technical area, it's a little tricky to define what the right technical skills are because it is less important that you have a set of particular skills than you have the ability to acquire new skills. You have to be constantly moving on." (IM) 235.

The following lengthy excerpt from an interview with a senior information manager further illustrates the variety of tasks and skills:

"Basic skills that are necessary for maintaining the hardware, someone that is good at system administration and maintaining the computer systems and networks, and upgrading them. Responding. There is a lot of responding to troubleshooting peoples problems with their computer. There is a whole element of training users to use the particular system that you setup and touse the software... And there is a whole level of administering the operating system, maintaining, security issues, adding users, installing software, upgrading the software, virus protection, doing routine backups on the computers and databases. There is a whole range of skills related to maintaining and developing a relational database...a database administrator. Then websites are just crucial for the way that people do science these days, and more communicating to a broader community. We put a lot of energy into developing our website and periodically major reorganizations and upgrades on the website …we spend a fair amount of time responding to requests for data, even though a lot of times people can access the data directly through the website. We have a commercially bought software for users browsing the database, but we still get requests by email and phone and people drop by our office and want some kind of help (both from within the group and outside). Sometimes we get even more involved in the analysis for a particular paper that is being written, help to generate the figures, actually perform some of the data reduction for people. We're significantly involved in inter-site information management activities. That is something that I do as part of my job. Some examples of that: I worked with some other people to organize information management workshops…I am on steering committee for the information managers' group, IMExec... I will be going to a planning meeting at the Network office to … extend one of the technologies that have been developed at the LTER network into biological field stations." (IM) 236.

The views of two newer information managers:

"It is crucial … to expect professional presentation in management when the skill sets are so widely varied of what you would expect from, I mean, let's face it, in many cases you would probably need a team of 4 or 5 people that have vastly different skills. You probably need a database manager, you need a graphic artist, or graphic designer, you need somebody that is going to know Perl or JAVA, or JAVA script, or whatever, and that is before we even talk about GIS. I mean in GIS you could have GIS personnel, image processing. So my point is that there is a wide variety of skill sets. And it is really unrealistic to expect one or two people in an LTER information managers role to have." (IM) 237.

> "I just can't do my job; I have to acquire the background and the skills to get the job done. Whereas if I were more a true manager, and I had a staff, I think I have the vision and I know where I want to go. If I had people to say, this is what I want, make it so. I think I could definitely fill that capacity as you know, information, as a true manager and not the person that puts the hat on to be the programmer or web site developer and so forth. So I think that, that is really hard for me." (IM) 238.

Information managers periodically poll their community of site representatives to gain perspective on the diversity of site solutions as a particular issue arises and is articulated. A recent survey revealed the diversity of site IM personnel staffing arrangements.

> "some sites are generally committed to information management but will not devote much of their resources. Some sites may only have a half time person where we have five different people working toward the effort. Some other sites have one person doing everything from fieldwork to data management to any number of things." (IM) 239.

> "So I really see that as one of the biggest problems that the sites are only funding like maybe a half, one person at half time, or one person at 3/4 time. And so what that leads to is kind of low salary positions as well a lot of times, and so what happens is they can't keep these higher skilled people." (IM) 240.

Some information managers who have few information management personnel at their sites describe creating networks and collaborations to expand capabilities on various expertises:

> "those types of things or questions about, you know, technical aspects that [co-worker] and I can't answer, luckily we are in a building with computer scientists. And I can go right next door and bug somebody. And I frequently have to make cookies in pay back" (IM) 241.

> "I mean we are developing closer and closer ties ... [with an IT department]. They do basically computer and networking support and that type of thing. ... many times I can tap on down for expertise and that type of thing." (IM) 242.

But there is a cost associated with collaboration. Participating in networks/collaboratories requires often unanticipated or unrecognized work and effort:

> "Whereas, networking things is much more difficult to throw your emotional energies into it. It would be fine if it is FedEx and you are trying to manage that kind of a system where a tangible thing, a package, is the be all and end all. In our case it's not a thing that is the be all end all, it is a set of ideas that you really want to get out and have people that you trust and admire believe what you have concluded." (PI) 243.

> "And the inter-site work is challenging in several ways, but in terms of proposal evaluation, I have seen, what has happened is, that ah, if you do it at one site it can be cheap, but if you do it at a bunch of sites, it's expensive." (PI) 244.

> "And it takes quite a bit of effort to really work together, to come to some sort of shared common understanding of how to represent simple things ..." (IM) 245.

4.2 Ongoing change and increasing expectations

The LTER views of information management have changed over the years.

> "I think in the past when it came to information management and GIS it was looked upon maybe as a necessary evil. It was not fully supported by PI's. And I think that was because, you know, they viewed it as really just money that they couldn't use for their own research. And I think there is a definite change in that…in that there is an understanding that National Science Foundation and LTER net are going to be more demanding as far as information management …and the other reason is I think that there is a understanding from the researchers themselves that, that it is a valid

way for them to share their data and their research." (IM) 246.

"People have come to accept information management's part …more than in the beginning. Now …they argue over the nature of the network information system, but early on they argue over why would we have an information manager at all." (IM) 247.

The work of information managers undergoes change influenced by an interplay of the evolution of the profession nationally and of its social organization within the LTER scientific community (cf. Abbott, 1998). The tendency in LTER has been from merely managing data to more heterogeneous and complex management of data, resources, technologies and people.

"For me, and I think for all the information managers, it is becoming more and more things are being asked for them to do. It used to be that we just managed data sets, played with data all the time. Now we manage personnel data, the bibliography data, the GIS databases..." (IM) 248.

"… it [my work] has changed a lot over the years and it is still changing" (IM) 249.

"My job has evolved a lot over time since I started ... at that time it wasn't a team. I did everything on my own. Gradually I moved to a more management position" (IM) 250.

"my own position has evolved over time … when I first started, I was sort of it. I was the only person on the information management staff. We did have people who were doing things like maintaining the computers, and doing a little bit of networking and that kind of stuff that we had. But I had to do all the database administration and handle the data requests and that kind of thing. And as our project has grown I have moved into a much more administrative kind of role, and [co-worker] who is working with me now, and he does most of the database administration and all of the sort of system administration stuff. And in some ways that can lead to me getting more out of touch with, actually being able to do the technology. So I am still involved in a lot of the design of the system, and the interactions with our researchers to decide how we want things to progress, and some of the database design and those kinds of things, but a lot of the more day to day stuff that I used to do I don't do anymore. Now I no longer program. When I first started I was doing some of the programming and stuff, and I don't do that anymore. And in the fall I am going to be going, almost full time, and I have always been less than full time and in recent years half time, and our lead PI wants me to spend most of that additional time doing research, which is fine with me. But there is a part of me that thinks, well gee, I should really be getting in there and learning XML, and some of the new tools that are coming on so that I would stay more connected with the nitty-gritty of the technology." (IM) 251.

Expectations of IM have changed with regard to all three relationship areas identified in the previous chapter between science, data and technology. This reflects changes that have taken place in the broader political, funding, and technology development contexts (e.g. NSF, NCEAS, ESA, domain and institutional partnerships).

Relationship: Science – IM

Network challenges prompt change in LTER site science work

"...that we are a network is different. There are scientific challenges that require collaboration across ecosystem types at much larger temporal and spatial scales. Part of the LTER mandate is to address these issues that couldn't be addressed by people collecting data just at one particular site. So, over time the components of building this network that will interact in a way to meet the scientific challenge have evolved too. And I think that the information management group has been a piece of how it has evolved at the network level. " (IM) 252.

"I thought that one of the important goals of the LTER program is to foster inter-site synthetic research, so I wanted to get a dialogue going between information managers and researchers reflecting on their process ..." (IM) 253.

Relationship: Data – IM

The expectations for data have changed from well-curated data to a focus on public availability and data accessibility. After the explosion of technology availability and of data and metadata accessibility as described in previous sections (2.1.1 and 3.3), questions of data discovery and queriability arise:

> "..right now we are working on doing a dynamic link between our website and our database so that people will be able to interact dynamically with the database through our website and do queries of the database, have the ability to bring up a subset of data that they are interested. That is an example where we have hired a computer science student with good programming skills to create a piece of software for us." (IM) 254.

The data challenge becomes even greater with recognition of the value of cross-site and cross-domain work that captures multiple dimensions of the sites:

> "The focus on, or the encouragement to do, cross-site synthetic work has been there for a long time. And there have been, some projects have been quite successful like the LIDET project. Especially experimental projects ... a lot of the discussion in the information managers' group is how to facilitate that electronically. That is, how can you, how can we take these resources…first how you make them available on-line, then how do you make them available in such a way that they can support cross-site research." (IM) 255.

Relationship: Technology – IM

The technological infrastructure has changed, e.g. from mainframes to personal computers, local networks, and the internet, presenting information managers with the challenges of technological diversification.

> "We develop the tools because we were taking a legacy data bank, and moving it from the old mainframes into a new system." (IM) 256.

> Technology has changed so much in the past 20 years. The beginnings of the LTER were at the very beginnings of the micro computer revolution. When I started, a lot of our activities were on a campus mainframe. Very little was done in terms of computers in our own lab. And now we are running our own machines; we do not use campus machines ... anymore." (IM) 257.

As tasks and skill sets have increased and diversified, also the responsibilities and appreciation for information management as well as information managers has grown.

> "There was a time when the data management end of the business was that, and it was not information management. The people who run the data and information management within individual LTERs are quite different in their backgrounds. They aren't trained computer scientists for the most part, some are. Some are Ph.D. people and some aren't, but they have a very good collegial working relationship. They have been highly successful. More and more sites, have information specialists who are PI's as opposed to staff. So that has been, and the executive committee has had a information management person on for at least the last 4 or 5 years and there will be another along when we have our replacement vote this year, I am pretty sure. That is a systematic acquisition in status, for a task area that we have always had in LTER. But they have come to have a much higher status in recent years. They have far larger role in decision making than in previous years." (PI) 258.

> "The resources allocated to information management and the appreciation for it have certainly grown. Early on there were sites that were putting very little of their resources into information management. Almost at the level of the secretary keeping track of the data, something really minimal. And it just became clear that those kinds of token efforts were not going to be acceptable, through the review and proposal processes and through the experience of the sites themselves. So there is just the norm that information management is an important part of each site's activities and part of what they are going to be evaluated on." (IM) 259.

> "They have a core of people who deal with the information management. And I think by and large they have certainly more respected and more paid attention to, maybe not as much as they would like, but certainly a great deal more than previously. And I don't know what I can attribute that to other than, other than its been dictated in dollars available and the recognition that has to be done for the legacy of data." (PI) 260.

4.3 LTER IM Committee: a forum for dialogue

The LTER Information Management Committee has provided a forum where information managers reflect on issues and have been able to take action. In 1989 an effort was made to list information managers' tasks (see Appendix 2). In 2001, the IM committee changed their job title from data manager to information manager, explicitly acknowledging the ongoing change processes and added responsibilities.

> "It was as if acknowledging the existing expectations and unfunded mandates, making visible complexities of work by changing the job title". (IM) 261.

> "I see the term information manager for a site as being because I know the value of it and so forth, I see it as being sort of prestigious. There are some folks that still perceive information management as this is the person who you hired to type the data into the computer." (IM) 262.

In addition to addressing the changes evident in information managers' work, the committee has also developed its ways of working to accommodate ongoing changes in the network.

> "we have these meetings every year and we engage in these activities that transcend the views and scientific agendas of site based science. So we've kind of evolved from recognizing that information managers needed to meet. And a lot of different kinds of things happen at those meetings, all of which support us functioning as a network and support excellent information management at each individual site. ... We are realizing that there is a need for more communication and coordination to take place so we have a steering committee. We didn't have a formal steering committee during the early years. During recent years we started to have the steering committee meet midway between the annual meetings of all information managers. And this year we realized that even that is not enough to keep us on track ... and people get energized and projects get started and then sometimes it is hard to sustain that over the course of the year if you don't have any other built-in contact. So this year we have started regular conference calls of the steering committee. And we spent a lot of time last summer as a steering committee identifying functions that needed to take place and then assigning those functions to various members of our group, … Who's on various working committees... will tend to get somewhat lost from year to year. We are trying to evolve our way of working more effectively ... because all of us also have our reponsibilities at our site too. " (IM) 263.

> "We are beginning to realize that because we have these big time commitments at out sites and network level stuff is added on, on top of that, at least for some of these activities, we have to begin to find sources of funding. We have been talking about for example a fellowship from the network office so that one of the information managers who has been working on a particular project might be able to go to the network office for a month and have their salary paid." (IM) 264.

> "Quite awhile ago, used to be that people that were kind of taking a leadership role. It was a very in-group kind of thing, and we moved to a system where people were elected, to serve terms and rotate off. And we also make an effort to get people that aren't on the Exec Committee involved in other activities, like being Databits editor ... And I think we should try to do more of that because people can do small pieces and there is a lot of talent out there, and it is a diverse group with different kinds of talent ..." (IM) 265.

Members of LTER community reflect on the Information Managers Committee's role within the network: e.g. from reactive to more proactive:

> "One of the things that struck me about the information management community in LTER, is that they have been extremely proactive, and very responsive to demands not only at the site but at the network level, in fact extraordinary responsive at the network level." (PI) 266.

> "It is important to the information manager to be proactive and to come up with their vision, where the site is going, how it connects to the network IM level and to provide that plan for the site." (PI) 267.

> "So I feel a responsibility that when I go to those meetings, I am not only speaking for me and my site, but I am also holding a place for everyone else." (IM) 268.

> "... because it was so important for us as a community to be recognized in not a custodial service role, but in a leadership role." (IM) 269.

4.4 Potential change processes

One LTER PI speaks of change as an interesting problem:

> "It is an interesting problem because we are in the midst of a system that is undergoing these sort of phases, and to even have this awareness of how it is, is quite difficult, isn't it? Kind of fun. (PI) 270.

… while another sees the long-term funding of LTER as a kind of change process in that it contributes to an unusual perspective:

> "But you know I think it's a program in ecological research that emphasizes, focuses on long term, particular secondary large scale research. Its built on the work done at individual sites which over time leads to the synoptic understanding of the sites where the work is done, multidisciplinary understanding of sites. And that it's built on the ability to develop comparisons across sites that lead to generalities how systems do or don't work. Because the funding is long term, the research is long term, so able to contribute in an unusual perspective on these systems." (PI) 271.

… yet another points out the change resulting from reviews

> "There was a shift in how the LTERs [see information management]. It became an expectation as part of the site visit and as part of the review process. We are also graded on information management and what we are doing. And so when it became an official part of the review process, and it was by, almost by edict then became more acceptable. And there has always been the tension of how much money do you put into information management versus how much money goes to the research project. (PI) 272.

The LTER 20 year review report (LTER, 2002) gives several ways to address IM and IT development including integrating in elements of research collaboratories and informatics. It reviewed the 10-year report recognition of a need for increased funding from NSF, particularly for research and data management at each of the sites. The recent LTER white paper (LTER 2000-2010, 2002) presents five goals that in effect initiate change as they prioritize and focus the community's next decade of research. Goal 2 is to support site synthesis with database development listed as a mechanism along with site volumes, network-wide synthesis projects, and multi-site synthesis projects.

This is the challenging setting in which information managers evolve their practices and develop technologies for ecological research. Processes may appear to overlap if evaluated solely by technological criteria:

> "Research scientists that are out there collecting data and writing papers and making graduate students, are not going to see the value of sitting around and spending a lot of money to develop computer software and algorithms and data search engines. They are not going to see the value of that. But once it works, they will begin to appreciate that." (PI) 273.

> "And there is still a big tension between those who want to design the network information system and those who just want to use the information effectively being collected…but I think we are on the verge, I think, of there may be a change. There is a growing interest in good information management that is coming from the ground-up at the sites that results from the need to manage the information that is already available." (PI) 274.

The fields of science and technology studies along with organizational and social informatics draw from sociology, anthropology, business, and history of science to provide alternatives for understanding change. The complexity of information systems together with collaborative science models suggest that such alternatives may be complementary. Rather than being mutually exclusive, they may help in triangulating in on an understanding of complex environments. Observations and interviews can add historical context, comparative studies, and multiple perspectives to participant reflections on potential change processes and paths along which IM development could proceed within the LTER network.

The process of change and the rate of change are sometimes noticed in retrospect or by contrast:

> "But there was at one time an effort to try to get each site to at least have some kind of an agreement as to what they would have and to agree on various software programs and so on. That turned out to be a futile effort. But it probably doesn't matter in hindsight, but it hurt then and hurt in-between, but now its almost at the point it doesn't matter because everybody is interoperable anyway. And if they are not interoperable, they know it, and they try to make something happen so they are. So I am not so sure that you could say that they changed, I think the technology just sort of overwhelmed them and now they have it. It isn't as they though they planned to make that change. There were people, voices in the wilderness saying, we should do this that or the other thing. But I think that change in technology was faster than the change in the human beings. So that kind of change kind of happened. (PI) 275.

> "And I would suppose that is one of the more fundamental pieces of change. Now, within our own group, some parts of our group are willing to take advances in technology and move with them instantly and others are not…So I think the dynamics of change because we are a multidisciplinary group, various elements of the discipline will have a different response time to change. And then some people are more conducive to change than others, that is just the nature of humans." (PI) 276.

4.4.1 Enabling inter-site research

The recent LTER white paper (LTER 2000-2010, 2002) discusses the 3rd decade of LTER as the 'decade of synthesis'. One of the five goals listed is to increase experimental and comparative cross-site research.

> "Modern information management has become crucial for inter-site ecological research. Any scientist who has gathered data from diverse data sources has dealt with issues of making data comparable, dealing with multiple data formats, structuring the aggregated data in a form that facilitates answering research questions, and providing access to the data and derived data and documents to colleagues. Without an adequate system for managing data, the magnitude of the effort needed for dealing with these issues for large and complex data sets can become a barrier to undertaking inter-site research." (IM) 277.

> "Well, what we have agreed to do now is that for every science theme that we come up with like the invasive species one, which is the first one we will do this. We will create a data management for that theme, so we work with the data managers right from the beginning to pull the data together to try to create, comparable, compatible datasets so we are in a position to look at all the data not just information, not interpretations and try to come up with an analytical capability to answer the question. And so that is new, but it has been driven by network level stuff, driven by the need to actually look at the data and use the data in synthetic activities. It also obviously brings to the forefront the value of the data and the value of the management of the data and now it is the real need which is making all these disparate ways of managing data compatible." (PI) 278.

> "Once the data were available in the database, researchers were able to readily select subsets of water bodies whose lengths of record and spatial locations were best suited to their particular question, thus maximizing their productivity. The availability of a database at the workshop supported the generation of important research questions as the scientists interacted." (IM) 279.

4.4.2 Information management research

At a research site, data management support for science traditionally has involved 'doing' a job, carrying out a prescribed task. Historically, a job well done is a job clearly defined so it can be performed efficiently, that is routinely and unnoticeably. In contrast, today an information management job may involve conceptual development and technology research. Evolving today are answers to the question: What is informatics research? In working on schema design, scenario development, legacy system update, metadata creation and multi-level interface coordination as well as prototype building, there is also the less well outlined work of implementation, testing and assessment that oftentimes requires innovative work.

> "The role of information manager at a site, that clearly there is this service role of sorts, and then the more research oriented." (IM) 280.

> "information managers … have this dual, on the one hand much of what they do, it supports the role of the site at the network. On the other hand, many information managers have research interests in informatics and would like to be doing cutting edge work as well." (IM) 281.

> "…we are all [information managers] recognizing that the building of the NIS [LTER Network Information System] is as much of a scientific process as it is a data management process." (IM) 282.

> "There are databases that are not even related to research data that facilitate operating as a network, like a personnel directory..an all site bibliography. A lot of work and discussion and evolution about how these things are going to be put together. And a lot of learning experiences about what things work, what things are difficult to maintain, where the actual information should reside ...at the site or collected in a central way) ... and how to keep things updated and all these kinds of issues ." (IM) 283.

Recognizing, integrating and legitimizing researcher role as part of information managers' identity and work requires both individual and community reflection, and establishes the need for training in research activities, such as writing and presenting papers.

> "With SCI2002 ...will the PI's have had anything else but annoyance out of this whole process? I decided to spend time doing it [the conference and paper] because it was important in terms of the learning environment and mentoring … Once it is done, it is gone, unless it has penetrated into the PI's' understanding that the time their data manager was sort of having a crisis of trying to submit a paper was not a diversion from important tasks at the site but rather was an adding to the learning of their information manager, of ultimately long term visibility of what that information manager could provide. That it was a publication that would go on the [site] publication list. And even that some of the information managers who are doing this as a first time thing would not think to package it that way. And they could simply feel bad that the PI's were having their service disrupted. So that would be a short-term view. But if we do not take the time to provide that longer term context, in some sense the short term will dominate..." (IM) 284.

> "I have been fortunate enough to participate in quite a few of the training workshops...a book volume...workshops at one of ESA meetings and then we had everbody write up small papers that were then published as a web volume." (IM) 285.

During the first decade of LTER and before the internet, there were printed catalogs of LTER datasets made available. In the second decade, accessibility of datasets was achieved after Internet infrastructure and web tools burst upon the scene. To address the third decade LTER goal of interoperable datasets, important IT and domain research is needed for developing concepts, technologies and processes for interoperability and synthesis.

4.4.3 Ecoinformatics

The LTER 20-year review states "The general term informatics has emerged as a way of referring to such applications of information technology, and of emphasizing their power in relation to the more traditional aspects of data management." The committee's recommendation #10 is to establish informatics as a core function.

Informatics is a term sometimes narrowed to refer to a specific discipline focus as in 'bioinformatics', 'geoinformatics', and 'ecoinformatics' (Michener et al., 2002).

> "Informatics is the study of computer science and information technology as it would be applied to a discipline to improve the and facilitate the scientific process in that discipline. So I use informatics broadly because it, you know, the stuff we do in ecology, is fairly crosscutting in terms of the science. So it involves everything from like I said ecosystems to community down to molecular types of research. The group, there is a group of people in the molecular community, that coined the term bio-informatics and that has been around for much longer. And they sort of, you could almost say they co-opted the term bio-informatics but they didn't include all the biology, they only included molecular biology. But they mean anything after the process of acquiring data that has to do with managing data, analyzing data, modeling data, mining data, doing theoretical sorts of simulations, stuff like that. So they apply it very broadly. And so we are really using that term in a similar vain. Except, so sometimes we call it ecoinformatics and sometimes we call it environmental informatics and sometimes we call it informatics and sometimes we call it bio-informatics and sometimes we call it information technology. But none of those terms are really significant, they are all basically the application of the technologies to facilitate the discipline and there are lots of ways, lots of areas in which that can happen." (IM) 286.

The ecological research network provides an important environment for informatics work. Ecological science grounds the work of information managers:

> "It's a long-term ecological research, not long-term informatics program." (IM) 287.

… yet information managers are stimulated by elements of informatics:

> "I have to have some aspect of what I do be information manager, or informatics research oriented, not just serving data management archive purposes. So I find a lot of stimulation and that's why I participate in this network ..." (IM) 288.

… and closely connected with science:

> "So I feel like those institutions are great for making major strides in the technology, but I think they might be too far removed from the science…"." (IM) 289.

> "For some people they want their one leg in ecology and one leg in computer science. But then you also have people that are really primarily computer scientists, and they don't really care whether or not its savings account that they are managing or the number of leaves on a tree that they are managing. But those people don't typically come to LTERs because they are not intrigued by that interchange of the ecology and the science with the IT." (IM) 290.

…and yet aware of their community role:

> "We were really pivotal in leading the community in recognizing the value of IT and IM. It is only recently that the community of ecologists has been more successful than just the computer scientists in going after those dollars." (IM) (SS/AND) 291.

4.4.4 Domain mediation

Information managers see that their experience as generalists and their immersed position in ecological research settings provides them with a special understanding needed in ecoinformatics development. Outside technology developers often are more abstract, with an emphasis on computational elegance and power rather than the as yet unarticulated everyday hurdles and the long-term science support:

"the LTER information management community, because of where it sits, sees most of the people who are doing, the specialists that are, you know, that are doing the big projects … they [the specialists] maybe run into a couple of ecologists a year, maybe. They are embedded in an environment of computer science and information technology. On the flipside, the LTER information manager in the LTER is an information manager embedded in a matrix of ecologists. And that gives them I think some special insights into what will work in their community and what won't. I mean over the years we have seen, you know, lots of big expensive computer science projects that have the best computer scientists there were doing the cutting edge stuff, that … essentially failed because they were not able to maintain contact with the community that the project was aimed at." (IM) 292.

"And the other thing that also we have seen over the years, is that, it is the unusual computer trained technician type, or you know computer expert, who can transition into this community and really be successful. Because we have a history of having sort of hot-shot computer experts jump in and then they are telling us all the stuff we are doing wrong. And their, oh no you gotta drop this model, you're acting like a bunch of bed and breakfasts. You gotta act like a hotel chain. We gotta centralize this stuff, we gotta get everybody on-line, we gotta start doing it the same way, you know. And I've got the vision, I've got the computer system that will allow us to do it. You know, boom, boom, boom. And generally speaking, I mean, we're willing to let, you know, I mean I am not saying that somebody can't do that, that they don't have good ideas, and I am more than willing to let them test them against the community. But almost inevitably what happens is within a year to two years, the person has moved on. Because working at an LTER site, they keep running into this impenetrable wall called ecologists. And you know, their contention is, 'OK, look, ecologists are a bunch of luddites ... they are just not forward looking, that they aren't.' I think they have completely missed the boat on that. I think that ecologists are more than happy and are hungry for things that will work, that will help them support their science. But they are not and in no way tolerant of doing a bunch of stuff that doesn't support the science, that doesn't create science products that they care about, that doesn't support the way that they work and think. And that's usually these, the folks are not coming in with that … there are some folks who you know, are able to adapt, that are able to make the bridge, but it is usually been easier for somebody whose training is more in ecology, to pick up the technology, than for somebody whose training is primarily technology to pick up the understanding of the way science works." (IM) 293.

"It [ecoinformatics] needs to be done by people that have some close contact with the problem, otherwise it is just database systems, with no applications. So I think that in defense of the people who say it's got nothing to do with ecology and therefore it doesn't belong, they are probably wrong because I think the science needs it, and it is going to take ecology informatics people to build it." (IM) 294.

"Susan had a slide where she talked about the critical role of mediators. And you know, the degree to which it worked had a lot to do with when you could get people who had a foot in both camps. ." (IM) 295.

There is a special character to IM work that stretches from traditional computer science and technology development to ecological and information sciences, that mediates between domains and specialists at the interfaces of science, data, and technology.

5 Continuing the Dialogue

This report aims to make visible some of the complexities involved in the work of information managers with its tacit and heterogeneous knowledge. Our ethnographic focus combines the ideas of technologically mediated work practice and participatory design with the concept of articulation work along with the notions of trajectory and the sociotechnical dynamics.

5.1 Communication occurs

Ecological information management is complex because of the heterogeneity of long-term data, the difficulty of identifying critical data, and the challenges of documenting datasets at a time of developing standards and emerging understandings about interoperability. There are a diversity of roles for an information manager in terms of providing support for heterogeneous data as well as for environmental science and technological infrastructure. Within the information management arena, we have discussed the ongoing work required to balance relationships and tensions. Each task, each role, and each balance involves communication whether with scientists and technicians, with classrooms and the public, or with technology and databases.

Because the informanagement role tends to multiple tasks, it may develop into one of the "hubs of communication" especially in federated endeavors

> "… the information manager is a hub for a lot of communications because they talk with everybody and they see the inconsistencies and non-alignments, they have the tools of technology to think about applying to some of that, and then they have the reality checks of what is really possible, having seen, ya know, doing it with the data, what is really possible." (IM) 296.

> "You have all the infrastructure and planning but you still need the champions, dedicated people and unfortunately it seems that corporate/institutional memory is trapped in the brains of those people, hard to get out." (IM) 297.

The frequency of unforeseen issues to handle and of technology developments to incorporate are factors that bring the need for continual review, assessment, rebalance and communication to information management work.

5.2 Considering the contexts

A trajectory, with a beginning and a direction, provides a conceptual framework within which to gather the multiple threads of information work over time. Our analytic approach uses the concept of 'articulation work', a research tradition in ethnography and social science (Maines, 1991). It is a concept used in analysis that highlights elements of everyday practice that may be lost in daily task oriented work where logistics and communications may be dismissed as trivial. Strauss and co-authors (Strauss et al., 1985) state that articulation work "must be done to assure that the staff's collective efforts add up to more than discrete and conflicting bits of accomplished work." In ecological terms, this is related to the concept of the whole being greater than the sum of the parts (Golley, 1993). In the technological realm, articulation assumes a variety of forms in the development process including the communications that occur during the requirements engineering work (Jirotka and Goguen, 1994). Reflexive or feedback system approaches take into account cyclic processes in a project lifecycle, rather than a linear one-time set of steps. Such approaches are an integral part of ethnomethodology e.g. protocol analysis (Goguen, 1994), participatory design (Blomberg, et al., 1993; Karasti, 2001) and narrative frameworks (Linde, 2001).

Star's (1991) insights on the sociology of the invisible in medical work are pertinent to the realm of ecological information management where data and ultimately knowledge bases are also highly visible entities in contrast with the largely invisible information management process. The observation that "Work is the link between the visible and the invisible" gains added impact with the understanding that "Articulation work is work that gets things back 'on track' in the face of the unexpected, and modifies action to accommodate unanticipated contingencies." It addresses the need for communications between the many agents involved in a complex work trajectory with multiple threads such as the vision and planning, the implementation and monitoring, in addition to the care work itself and task coordination. Within the LTER, the complexity of interdependent tasks stretch from sampling in the field through institutional and individual offices and into the digital domain. Unexpected events, contingencies that arise locally, create venues that are not fully rationalized and hence lack an associated 'Standard Operating Procedure' or a 'best practices'.

5.2.1 IM Role

The LTER information managers work at the intersection of three sciences: environmental, information, and social sciences. One view of this environment is shown in Figure 4 (from Baker et al., 2002): environmental science with earth system drivers; organizations with human practitioners, cooperative teams, institutions and networks; and information science with long-term data care, federated systems, and technology drivers.

An informal network of individuals who share a common interest that cuts across traditional organizational departments is known as a community of practice (Lave and Wenger, 1990). Communities of Practice (CoP) may develop within organizations, across institutions, within networks, or at the intersections of domain sciences. In Figure 4 information management, concerned with the gathering, storage, and access to data at the intersection of environmental science and information science, appears as a community of practice at the overlap of Environmental and Information Sciences; Computer Mediated Communication (CMC) appears at the interface of Environmental and Social Sciences; and Human Computer Interaction (HCI), Computer Supported Cooperative Work (CSCW), Social Informatics (SI) and Participatory Design (PD) are at the intersection of Information and Social Science boundaries. The Center (U) designates a union of understanding of the three science domains.

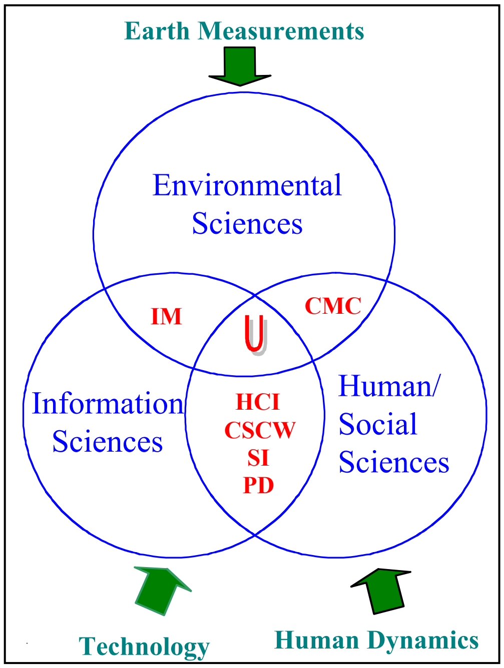


Figure 4: Domains and Communities of Practice (CoP) (from Baker et al, 2002)

The number of CoPs shown in Figure 4 at the join between Information Sciences and Social Sciences is an indicator of the complexity at the boundary. Social Informatics refers to the body of research and study that examines social aspects of computerization, including the roles of information technology in social and organizational change and the ways that the social organization of information technologies are influenced by social forces and practices (http://www.slis.indiana.edu/si/).

Computer Supported Cooperative Work is a term to describe the understanding of the way people work in groups with the enabling technologies of computer networking, and associated hardware, software, services and techniques. It is concerned with designing shared information spaces and supporting heterogeneous, open information environments that integrate existing single-user applications (http://www.telekooperation.de/cscw).

Participatory Design is an approach to the design and development of technological and organizational systems that places a premium on the active involvement of workplace practitioners in design and decision-making processes (http://www.cpsr.org/program/workplace/PD.html). Participatory design work focuses on understanding how work is done and how technology can provide support for it.

A heterogeneity of data and knowledges creates an environment within which a multitude of interdependent decisions occur. With this understanding comes the notion of the information manager as a **change agent**.

> "But you know we, and it will be interesting because there are some folks who you know, are able to adapt, that are able to make the bridge, but its usually been easier for somebody whose training is more in ecology to pick up the technology than for somebody whose training is primarily technology to pick up the understanding of the way science works. And the diversity of the ways in which it works, that is the other thing, is that you know, there's no two investigators at my site that work exactly the same way. (IM) 298.

Technology is often viewed as a change agent, but we would add that it is the enactment, the integration and use of technology, that brings about change. Further, efforts to initiate change are often not aligned with the status quo so are frequently resource poor though perhaps innovation rich.

5.2.2 Design

We use the term 'information system' to describe a group of interdependent technical and social elements of communication and computation used within an organization. There is a valuable tradition within information system design of taking into account practitioners. This simple, straightforward concept that enables everyday practices is deceptively complex and often lost when projects scale into separate functional layers. The notion is sometimes folded into waterfall and life cycle models through use of the terms such as 'iterative' or 'feedback'; it is integral to engineering requirements and user-centered design.

> "When it came time to design this workshop, I thought, you know, these people don't want to talk, they want to just jump right into it ... actually getting things done. I was just totally surprised that people wanted to talk and talk and talk on the first day. The break-up groups didn't want to come back from break-up, they wanted to continue talking about EML ..." (IM) 299.

> "I think that the design that got put forth with the ClimDB project was I think a good design to improve the outcome. I hope it works because it places more participation. It wasn't just sending something to SDSC and saying here is what we need, build it. ..." (IM) 300.

> "Most of those projects are all designed, the databases are all designed here in the lab. The entry forms are designed here in the lab ..." (IM) 301.

> "I think [what makes the network work] is the people like, it is the collaborations between people. I don't think it is the design, there is no design for network" (IM) 302.

> I don't think it would fly at the coordinating committee if somebody said well all of you people doing biogeochemistry data, you're going to have to do this. So I think that there it is less of a network there; it is more of a community. It is a network in the sense that the money is being allocated through a network and the Network Office is providing some coordination, but I don't think they are directing research in the same way that they seem to be wanting to direct data management. (IM) 303.

5.2.3 Sociotechnical Systems Design

Data systems store and handle data while more complex information systems involve metadata and automated response. Both involve forward engineering and information technology engineering requirements design. As an alternative to optimized technical and cybernetic approaches to design, sociotechnical design principles include the use of work groups (much like LTER committees). Research is ongoing in many fields, from engineering to management science, from computer to information science. Cherns in 1987 suggested one early summary of principles of *sociotechnical* design:

- Compatibility
- Minimal Critical Specification
- Variance Control
- Boundary Location
- Information Flow
- Power and Authority
- Multifunctional Principle
- Support Congruence
- Transitional Organization
- Incompletion

A sociotechnical system is one taking into account both human and technical elements of a system and incorporating the view that local, everyday practices often represent unique points of innovation and knowledge making, of learning and understanding.

> "What is the network? In some sense it is a social organization ..." (PI) 304.

With sociotechnical perspectives providing integrative foundations from which to draw, the use of 'sociotechnical' as an adjective provides an informal, flexible term broadening the focus from technical to encompass more of the multiple facets of information systems or activities, the human and the technical and the relations between them.

5.2.4 Infrastructuring

For collaborative information systems as well as for collaborative science, the term 'infrastructure' appears frequently in the thoughts, plans, and conversations of LTER participants (see Quote #196). Reference may be to the technical infrastructure

> I am still UNIX oriented because it represents the common sharing infrastructure for our data and databases and for our group communication. (IM) 305.

> "So a great deal of my time and effort ... has gone into building up the computer infrastructure and network infrastructure that we needed to function as a department for the university and to function as an LTER site." (IM) 306.

> "If that is what you mean by infrastructure: capital, capital equipment. There are for example, line items in the NSF budget to buy very large pieces of equipment for something called a major research equipment fund." (IM) 307.

as distinct from the intellectual resources

> I would view, that, the total infrastructure resources that LTER has as being sort of less overwhelming than is the intellectual resources which are far more transferable". (PI) 308.

Perhaps the blending of social and technical infrastructures can be captured by today's emerging cyberinfrastructure vocabulary (Atkins, 2003). The LTER community has worked with social and technical infrastructure concepts, recognizing both intellectual and equipment capitals, since early on:

> "... well from the beginning the Network Office has provided a level of information and infrastructure that has been very helpful, particularly in the early days when email didn't work very well." (IM) 309.

Though often used as a metaphor to elicit thoughts of wires and intertwined channels of information systems, Star and Bowker (2002) challenge us to imagine the term 'infrastructure' as an active process taking in a full range of sociotechnical dynamics (to infrastructure) rather than as a structure to be completed (the infrastructure).

"Organizationally information managers are in the process of building new infrastructure to support the collaboration." (IM) 310.

"From the beginnings of the vision of there being a NIS, there has always been the idea that we should be working toward providing the information infrastructure that will facilitate operating as a network." (IM) 311.

"... from our past experience, what kinds of things have lead to successful collaboration between information managers and researchers in pulling together inter-site kinds of projects. And also to begin to look at how we might build some new information management infrastructure that would support that process, to look at the various components of putting together an inter-site research."(IM) 312.

I see that we are given a lot of, there is a lot of things we would like to do in information management, but they are kind of like unfunded. These unfunded mandates. You will put data on the web in two years after collection, and yet they haven't paid for the infrastructure to really do that. (IM) 313.

"As a community the LTER may be considered a social system or even a cognitive ecosystem with its own unique infrastructure. The success of such a community system depends upon whether important issues can be recognized, communicated, and addressed. In particular its critical for planners to keep in mind the differences between those who work to support the system and those who benefit from it." (IM) 314.

"I think criteria for which sites are reviewed have changed over time. Definitely. I think originally it was all based on how good the science was perceived, and I think now it really is built on, from this little short synopsis that I saw, on the overall strength of the infrastructure, your partnerships with like other groups." (IM) 315.

I can sort of see that reflected in the current distribution of infrastructures in LTER that sites ... that had really strong involvement of PI's and data management have really the strongest uh, infrastructures for supporting data sharing and information management and they tend to be the leaders in trying to push the network further and further as well. (IM) 316.

5.3 To be continued …

5.3.1 IM Work: Seeing the Invisible

The nontrivial nature of the informatics task is recognized by community participants such as this information manager

> "Knowing how to prioritize some of these things is really difficult. And how to express what seems trivial, to make it into something that it in fact brings cohesion and identity." (IM) 317.

and an LTER PI describing the coordination effort involved in gathering data together:

> "It's hard to, yeah, it's just one of these things that keeps goin' through your fingers like sand. You know, you think you can pick it up but as soon as you spread your fingers, it's gone. And you gotta pick it up again. I've noticed that on some projects, they get done multiple times, even by the same people because for some reason it refuses to be finished. I don't know why. But there are certain projects like that, or activities, that they just, either get out of date too quick, or there are so many things to combine up …getting them all correct, all the time turns out to be non-trivial." (PI) 318.

An LTER information manager sees LTER as a model with key elements:

> "And I think LTER is something that is more… they are aware of what sharing brings into the scenario, the need of sharing and, you know, just having the perspective of a long term research. Those two things are the key of LTER… And I think it is a model…. It is a model that not only for other scientists, it is a model for the rest of humanity ... to learn how to share in spite of the, you know, difference in culture, in perspective, in languages." (IM) 319.

So the work that goes into collecting and maintaining the data is largely invisible. The infrastructure work is often seen as mundane, boring and not worthy of that much attention.

> "people forget when they are at that level of intellectual, you know, that the little process, and the day-to-day processes are really

> important. And are really the backbone, the basis." (IM) 320.

> "So maybe that is the main thing, I think there is an important obligation to maintain the continuity of these datasets over the long term, and that means there are certain things that you just have to keep doing year after year, and that takes a strong sense of obligation. Some people get tired of doing something after a few years, and you have to realize, you know, during any 2 or 3 or 4 or 5 year stretch, the data may be pretty tedious and it may be kind of boring to do it but then you have to realize that what your are really contributing to is a 10 or 20 or a 50 year dataset and that some day somebody is going to look back on these datasets and realize how valuable they are, even if those of us are working on the program right now are not around to see that, I think we can take satisfaction and realize that we are building part of a much larger structure. So you have to have a certain amount of vision and faith in the future." (PI) 321.

Indeed, Bowker (2000) comments on the heroic nature of this work:

> "The information collection effort that is being mounted worldwide is indeed heroic".

5.3.2 Mutual learning: PD and our project

Participatory approaches consider design as a process in which technology users are participants whose practices are taken into account. Articulation and participation are a recognized part of the design process which may include iterative design and feedback systems. Participatory Design (PD), relevant to a range of development and maintenance processes, includes the use of observer-participants and interventions.

Intending to bridge qualitative research and practice using qualitative research methods for study, this project draws on Participatory Design. The LTER network provides a long-term case study in a real world setting of scientific cooperation and technology appropriation. Our project was conceived as a means of articulating and bridging the intertwined social and technical elements present in this. The LTER presents a complex research environment with a broad common focus on the importance of long-term data in understanding the earth ecologically, an acknowledged approach respecting both diversity and community, and an unusual long-term funding support. This research community operates at multiple levels so this report extends outside the more traditional, manageable small group inquiries of cooperative work. The multiple levels include multidisciplinary scientists cooperating on interdisciplinary site science (network of scientists), multiple sites cooperating on cross-site themes and federated information management (network of sites) as well as the network cooperating with other networks (networks of networks). addresses many of the difficult issues of cooperative research.

In keeping with the practices of participatory design where feed-back and dialogue are an integral part of the process, this comprehensive report was planned early on. Our intent is to provide the LTER community with an example of a qualitative research approach using ethnographic methods to explore their work and with an analytic overview of the 2002 fieldwork done in collaboration with LTER participants. Materials from talks and reports summarized below are included in the appendices.

Our dialogue with the community began in 2001. The project was introduced to the community prior to our 2002 fieldwork in an LTER Newsletter article that describes the NSF funded one year project (Baker et al., 2002). The dialogue continued throughout 2002 with presentations (February Information Managers Executive Committee meeting, August Information Managers annual meeting), newsletter contributions (Spring and Fall Databits), interviews and research site visits. Our analysis continues with students mentored by Helena Karasti using LTER materials Appendix 8.11; Karasti and Syrjanen, 2004), with contributions to the newsletters introducing notions of sociotechnical and infrastructuring, and with an NSF Human Social Dynamics Comparative Interoperability Study (Bowker and Baker).

We summarize our research and results both for selected social science communities (CSCW02, ECSCW03, HICSS04) as well as LTER community participants (poster, All Scientists Meeting 2003; handout, Information Manager's Meeting 2002, 2003). The Computer-Supported Cooperative Work conference workshop on metadata titled "Storytelling and Collaborative Activities" afforded the opportunity to explore the LTER ethnographic materials with a contemporary approach to metadata (Karasti et al., 2002). Because the traditional focus within CSCW has been on forms of work other than scientific collaboration, we chaired a workshop at the European Computer Supported Cooperative Work Conference this year titled "Computer Supported Scientific Collaboration (CSSC)" to provide a forum for discussions on communication and collaboration through software, databases, and infrastructure. A paper for the Hawaii International Conference for Systems Science focused on the ramifications of long-term for information management (Karasti and Baker, 2004).

5.3.3 Articulating the Issues

In today's federated workplaces, the concept of the 'cumulative mess trajectory' (Strauss et al., 1985; Star, 1991) provides a hint of the consequences of ignoring or proceeding informally with articulation in today's federated workplaces. A silence at any of the management or coordination levels, regarding processes composed of large amounts of invisible work, ripples through the multitude of subtasks within trajectories of organizational partnerships. Articulation is a critical strategy in the scaled, federated partnerships crafted to support the scaling of contemporary, research science. The complexity of issues in information management today brings new requirements to communities for dialogue to ensure development of design processes that can integrate workflow and contribute to effective infrastructure (see Quotes #242 and #243).

In order to assure the flow of information between the scientists and the information managers within LTER, the LTER Executive Committee has in recent years included an information management representative while the information managers have annually invited guests to their annual meeting including LTER scientists. As issues such as metadata and standards have come to the forefront stimulated by several NSF funded research efforts with community members and affiliates (Jones et al., 2001; McCartney and Jones, 2002), there is recognition of the need to formalize and add continuity to communication between information managers and scientists. The Network Information System (NIS) Committee was established in 2002 with members from both the information management and the scientist communities. The committee adds continuity to communications and builds common vocabulary through joint preparation of reports and presentations.

> "... but the information people are on the other hand very technology oriented, they often speak in unintelligible acronyms, honest to god. ... and I think actually they have done pretty well, they have realized that they have to get real examples from the scientists. Because, otherwise they are kind of talking about arbitrary terms, or abstract terms, and it just doesn't make any sense ... Ah, the unfortunate thing is roping scientists into participating, it's been really hard. Again because the motivations are just completely

> different. ... The scientists are actually trading off multiple things ...” (PI) 322.

> “... the problem with NIS in the LTER system where I think that there is a case of a clear misconception between what can be built and what the expectations are, and remedy was to propose to build this committee that consists of PI’s and data managers that together define the goals, and therefore hopefully share a more realistic common set of expectations. But it was sort of a reminder that there are these expectations and what you actually do, they can easily drift quite apart.” (IM) 323.

5.3.4 Potential forums of IM research

Multiple domains, multiple communities

Figure 1 presents the notion of the intertwined mundane and complex knowledge work of information management while Figures 2 and 3 provide ways to view information management work. Figure 4, representing the three domains of environmental science, social science and information science, provides a context for relating key concepts, literature and boundary areas important to the LTER community, ecological information management and long-term infrastructuring.

Journals, conferences, and publications

The number of domains and communities shown in Figure 4 result in an increase in the numbers and types of arenas available for information exchange. In addition, new forums are emerging that provide opportunities for publication in journals with broader scopes and in digital collections that accept databases for publication.

> “... as time goes on and people publish some very insightful papers, for instance there is the relatively new Springer Verlag journal Ecosystems, you know, that wouldn’t have been, had that big of a market 15 years ago. I think it is in its 5th year, and it has virtually taken off.” (PI) 324.

The rapid growth in the number of publication forums for information management and information systems is somewhat bewildering. A collection of some of these forums and summaries of the forms appears in the appendices. Attempts to provide an overview has taken the form of online priority rankings. With the proliferation in journal rankings, there are even summaries of the summaries which are outdated rapidly as new journals emerge Appendix 8.15). For instance, recently the *Journal of Information Technology: Organization, Management, Information and Systems* re-launched itself with a focus on ‘technochange’ and the *European Journal of Information Systems* was initiated to provide a global forum for theory and practice of information systems.

In addition to journals, there are a wide variety of conference venues, some established and some fairly recently organized. These meetings provide a mechanism for continued articulation and learning.

> “So to get involved in that social mix you have to go to a national meeting, you have to put yourself in a position that you can benefit from those things, other than just reading journal articles. So there is a real social function to these. “ (PI) 325.

Many of these conferences produce a proceedings which is another, sometimes rapid, avenue for publication in addition to journals with more traditional publishers. The LTER information managers as a group have participated as a community periodically in conferences, i.e.: EcoInforma in 1996 and Sixth World Multiconference on Systemics, Cybernetics and Informatics in 2002 (see Appendix 8.3).

> “If you start to publish things and produce your own, and go to like this computer conference and present, that is something [the site PI’s] can count in their attainment, and so the more of that type of thing you do, you hate to spend all that time, at least I do. I think it is

> a great thing. You should write it up, the good things that you have done, but it is just one more thing to do. But I think it has been really important." (IM) 326.

> "With SCI2002 when we said well for the impact of all the work we are doing, we are writing these papers, they'll be in the proceedings, on CD, we'll all know, we'll be talking to each other at that session, and other people." (IM) 327.

For this LTER information manager, participating in a conference prompts reflection and articulation about the work and the support of the work.

> For the SCI2002 I am thinking whether there is a way to package this instead of once again launching into working ... and not having support to do it. (IM) 328.

Hosting workshop sessions at meetings has provided a forum for the group to articulate ongoing work, to interact with other professionals with similar interests, to gain experience in professional communication formats, and to mentor colleagues within the LTER information management committee.

5.3.5 Closing with an opening remark

An early LTER data management report (Appendix 8.2) captures some of the lived, often unspoken underpinings, of the LTER information management community that continue today: full participation ('Data Managers from all seventeen sites attended..."), science-information management interface ('...all Principal Investigators ... discuss their expectations of dadta management with their data manager..'), and reflection ('assess them with regard to their own expectations and their perception of the expectations of the NSF...').

An information manager new to the LTER community expressed the potential and spirit of information management today:

> "But I think this is the infancy in some ways of information management. I feel very proud and very privileged to be in the infancy …I think at a certain level the people that are there now are very much the principal architects or innovators that are going to set the tone for the next however many years." (IM) 329.

Today's concern with long-term perspectives in concert with the growth of distributed digital technologies and scientific partnerships, creates a changing arena for scientific research. Within this arena, the role of information management is a work in progress.

6 Acknowledgements

We offer special thanks to the LTER community for contributing their insights and taking the time to share their professional experiences as well as to our co team member Geoffrey Bowker. This work is made possible by agencies supporting research in the understudied areas of infrastructuring, sociotechnical dynamics and collaboration science (NSF BDEI BDEI-01-31958; Palmer LTER grants OPP-96-32763 and OPP-02-17282, SBE/SES & OPP-04-05069, and HSD/SBE SES-04-33369), Academy of Finland; Finnish Cultural Foundation, and Scripps Institution of Oceanography).

7 References

Abbot, A. 1998. *The System of Professions: An Essay on the Division of Expert Labor*. Chicago, Ill., The University of Chicago Press.

Atkins, D. E., K. K. Droegemeier, S. I. Feldman, H. Garcia-Molina, M. L. Klein, D. G. Messerschmitt, P. Messina, J. P. Ostriker, and M. H. Wright. 2003. Revolutionizing Science and Engineering through Cyber Infrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, January.

Atkinson, P., and M. Hammersley. 1994. Ethnography and Participant Observation. Pages 248-261 in N. K. Denzin and Y. S. Linbcoln, eds. *Handbook of Qualitative Research*. State Publications, Inc., Thousand Oaks, CA.

Baker, K. S. 1996. Development of Palmer Long-Term Ecological Research Information Management. *Proceedings of Eco-Informa Workshop, Global Networks for Environmental Information, 4-7 November 1996, Lake Buena Vista, FL* 11: 725-730.

Baker, K. S., B. J. Benson, D. L. Henshaw, D. Blodgett, J. H. Porter, and S. G. Stafford. 2000. Evolution of a multisite network information system: the LTER information management paradigm. *BioScience* 50: 963-978.

Baker, K. S., G. Bowker, and H. Karasti. 2002. 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.

Baker, K. S., J. Brunt, and D. Blankman. 2002. Organizational informatics: site description directories for research networks. Pages 355-360 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.

Becker, H. S. 1998. *Tricks of the trade: how to think about your research while you're doing it*. University of Chicago Press, Chicago.

Benson, B. J. 1996. The North Temperate Lakes Research Information Management System. Pages 719-724. *Proceedings of Eco-Informa Workshop, Global Networks for Environmental Information, 4-7 November, '96, Lake Buena Vista, FL.*

Benson, B. J., and R. J. Olson. 2002. Conducting cross-site studies: lessons learned from partnerships between scientists and information managers. *Bulletin of the Ecological Society of America* 83: 198-200.

Berners-Lee, T., J. Hendler, and O. Lassila. 2001. The semantic web. *Scientific American*.

Blomberg, J., J. Giacomi, A. Mosher, and P. Swenton-Wall. 1993. Ethnographic field methods and their relation to design. Pages 123-155 in D. Schuler and A. Namioka, eds. *Participatory design: principles and practices*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Blomberg, J., L. Suchman, and R. H. Trigg. 1996. Reflections on a Work-Oriented Design Project. *Human-Computer Interaction* 11: 237-265.

—. 1997. Reflections on a work oriented design project. Pages 189-215 in G.

C. Bowker, S. L. Star, W. Turner, and L. Gasser, eds. *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. Lawrence Erlbaum Associates, Mahwah, NJ.

Bowker, G. C. 2000. Biodiversity Datadiversity. *Social Studies of Science* 30: 643-683.

Brand, S. 1994. *How Buildings learn*. Penguin Books, New York.

Briggs, J. M., and H. Su. 1994. Development and refinement of the Konza Prairie LTER research information management program. Pages 87-100 in W. K. Michener, J. W. Brunt, and S. G. Stafford, eds. *Environmental Information Management and Analysis: Ecosystem to Global Scales*.

Brown, J. H. 1994. Grand challenges in scaling up environmental research. Pages 21-26 in W. K. Michener, J. W. Brunt, and S. G. Stafford, eds. *Environmental Information Management and Analysis : Ecosystem to Global Scales*.

Brunt, J. W. 1998. The LTER network information system: a framework for ecological information management. Pages 435-440 in C. Aguiree-Bravo and C. R. Franco, eds. *Proceedings(RMRS-P-12) of North American Science Symposium - Towards a Unified Framework for Forest Ecosystem Monitoring and Research, 1-6 November 1998, Guadalajara, Jalisco, Mexico*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Brunt, J. W., P. McCartney, K. S. Baker, and S. Stafford. 2002. The future of ecoinformatics in Long Term Ecological Research. Pages 367-372 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.

Brunt, J. W., and R. W. Nottrott. 1996. The LTER network information system for the 21st century. *Proceedings of Eco-Informa Workshop, Global Networks for Environmental Information, 4-7 November 1996, Lake Buena Vista, FL* 10: 104-104.

Brunt, J. W., J. Porter, and R. Nottrott. 1990. Internet connectivity in the LTER: assessment and recommendation report. LTER Network Office, University of Washington, College of Forest Resources AR-10, Seattle, Washington.

Bødker, S., K. Grønbæk, et al. 1993. Cooperative Design: Techniques and Experiences From the Scandinavian Scene. In: *Participatory Design: Principles and Practices*. D. Schuler and A. Namioka. Hillsdale, NJ, Lawrence Erlbaum Associates: 157-175.

Callahan, J. T. 1984. Long-term ecological research. *BioScience* 34: 363-367.

Carson, R. L. 1950. *The Sea Around Us*. Oxford University Press, Oxford.

Cherns, A. 1987. Principles of sociotechnical design revisited. *Human Relations* 40: 153-162.

Davenport, T. H. 1997. *Information ecology: mastering the information and knowledge environment*. Oxford University Press, New York.

Coghlan, D. and T. Brannick 2001. *Doing Action Research in Your Own Organization*. London, SAGE Publications.

Franklin, J., C. Bledsoe, and J. Callahan. 1990. Contributions of the Long-Term Ecological Research Program: an expanded network of scientists, sites, and programs can provide crucial comparative analyses. *BioScience* 40: 509-523.

Goguen, J. A. 1994. Requirements engineering as the reconciliation of social and technical issues. Pages 165-199 in M. Jirotka and J. Goguen, eds. *Requirements engineering: social and technical issues*. Academic Press, London.

Golley, F. 1993. *A history of the ecosystem concept in ecology: More than the sum of the parts*. Yale University Press, New Haven.

Greenbaum, J., and M. Kyng. 1991. *Design at work*. Lawrence Erlbaum, Hillsdale.

Hanseth, O. and E. Monteiro 1996. Developing Information Infrastructure: The Tension between Standardization and Flexibility. *Science, Technology & Human Values* 21(4): 407-426.

Heath, D. 1997. Bodies, Antibodies, and Modest Interventions. In: *Cyborgs and Citadels: Anthropological Interventions in Emerging Sciences and Technologies*. G. L. Downey and J. Dumit, SAR press: 67-82.

Henshaw, D. L., G. Spycher, and S. M. Remillard. 2002. Transition from a Legacy Databank to an Integrated Ecological Information System. Pages 373-378 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.

Hobbie, J. E. 2003. Scientific accomplishment of the Long-Term Ecological Research program: an introducton. *BioScience* 53: 17-20.

Hobbie, J. E., S. R. Carpenter, N. B. Grimm, J. R. Gosz, and T. R. Seastedt. 2003. The US Long-Term Ecological Research Program. *BioScience* 53: 21-32.

Holstein, J. A., and J. F. Gugrium. 1995. *The Active Interview*. Sage Publications, Inc., Thousand Oaks.

Iivari, J., and K. Lyytinen. 1998. Research on Information Systems Development in Scandinavia-Unity in Plurality. *Scandinavian Journal of Information Systems* 10: 136-185.

Ingersoll, R. C., T. R. Seastedt, and M. Hartman. 1997. A model information management system for ecological research. *BioScience* 47: 310-316.

Jirotka, M., and J. Goguen. 1994. *Requirements Engineering: Social and Technical Issues*. Academic Press, London.

Jones, M. B., C. Berkley, J. Bojilova, and M. Schildhauer. 2001. Managing scientific metadata. *IEEE Internet Computing* 5: 59-68.

Karasti, H. 2001. Increasing sensitivity towards everyday work practice in system design. Pages 151. University of Oulu, Oulu University Press, Finland.

Karasti, H., and K. S. Baker. 2004. Infrastructuring for the long-term: ecological information management. *Proceedings of the Hawai'i International Conference on System Sciences (HICSS) 2004, 5-8 January, Big Island, Hawaii*. IEEE, New Brunswick, NJ.

Karasti, H., K. S. Baker, and G. C. Bowker. 2003. Ecological Storytelling and Collaborative Scientific Activities (in press). *SIGGROUP Bulletin*.

Karasti, H. and A-L. Syrjanen. 2004. Artful Infrastructuring in Two Cases of Community PD. Proceddings of the Participatory Design Conference, Toronto.

Kling, R. 1999. What is Social Informatics

and Why Does it Matter? *D-Lib Magazine*: 1-23.
Lave, J., and E. Wenger. 1991. Situated Learning: Legitimate Peripheral Participation. Pages 1-38. Cambridge University Press.
Linde, C. 2001. Narrative and social tacit knowledge. *Journal of Knowledge Management* 5: 160-171.
LTER. 2002. LTER 2000-2010: a decade of synthesis, http://www.lternet.edu/archives/documents/presentations/lter_2010/.
LTER Twenty-Year Review. 2002. http://intranet.lternet.edu/archivs/documents/reports/20_yr_review/.
Magnusen, J. J. 1990. The invisible present. *BioScience* 40: 495-501.
Maines, D. R. 1991. *Social Organization and Social Process: Essays in honor of Anselm Strauss*. Adline De Gruyter, New York.
McCartney, P. H., and M. B. Jones. 2002. Using XML-encoded Metadata as a Basis for Advanced Information Systems for Ecological Research. Pages 379-384 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.
Melendez-Colom, E. C., and K. S. Baker. 2002. Common information management framework: in practice. Pages 385-389 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.
Michener, W. K. 1986. Data management and long-term ecological research. Pages 1-8 in W. K. Michener, ed. *Research Data Management in the Ecological Sciences*. University of South Carolina Press, Columbia, SC.
Michener, W. K., and J. W. Brunt. 2000. *Ecological Data: Design, Management, and Processing*. Blackwell Science.
Michener, W. K., J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford. 1997. Nongeospatial metadata for the ecological sciences. *Ecological Applications* 7: 330-342.
Michener, W. K., J. W. Brunt, and S. Stafford, eds. 1994. *Environmental Information Management and Analysis: Ecosystem to Global Scales*. Taylor & Francis, London.
Michener, W. K., J. W. Brunt, and K. L. Vanderbilt. 2002. Ecological Informatics: a Long-Term Ecological Research Perspective. Pages 390-395 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.
Michener, W. K., J. H. Porter, and S. G. Stafford. 1998. Data and information management in the ecological sciences (DIMES): A Resource Guide. *Proceedings of a workshop held at University of New Mexico, Albuquerque, NM, 8-9 August, 1997*.
Odum, E. P. 1953. *Fundamentals of Ecology*. International Thomson Publishing, Oxford.
Olson, R. J., J. M. Briggs, J. H. Porter, G. R. Mah, and S. G. Stafford. 1999. Managing data from multiple disciplines, scales, and sites to support synthesis and modeling. *Remote Sensing of Environment* 79: 99-107.
Olson, R. J., and R. A. McCord. 1998. Data Archival. Pages 53-58 Data and

information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque, NM.

Olson, R. J., L. D. Voorhees, J. M. Field, and M. J. Gentry. 1996. Packaging and distributing ecological data from multisite studies. *Proceedings of the Eco-Informa Workshop, Global Networks for Environmental Information, Lake Buena Vista, FL* 10: 93-102.

Plowman, L., Y. Rogers, et al. 1995. What Are Workplace Studies For? In: *The Fourth European Conference on Computer-Supported Cooperative Work (ECSCW'95)*, 10-14 September, 1995; Stockholm, Sweden, Dordrecht: Kluwer Academic Publishers. 309-324.

Polland, B.D., 1995. Transcription Quality as an Aspect of Rigor in Qualitative Research. Qualitative Inquiry 1(3): 290-310.

Porter, J. H., and J. T. Callahan. 1994. Circumventing a dilemma: Historical approaches to data sharing in ecological research. Pages 193-202 in W. K. Michener, J. W. Brunt, and S. G. Stafford, eds. *Environmental Information Management and Analysis: Ecosystem to Global Scales*.

Porter, J. H., B. Hayden, and D. Richardson. 1996. Data and information management at the Virginia Coast Reserve Long-Term Ecological Research Site. Pages 731-736. *Proceedings of Eco-Informa Workshop, Global Networks for Environmental Information, 4-7 November '96, Lake Buena Vista, FL*. Environmental Research Institute of Michigan (ERIM), Ann Arbor, MI.

Porter, J. H., and K. W. Ramsey. 2002. Integrating Ecological Data: Tools and Techniques. Pages 396-401 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.

Reason, P. and H. Bradbury, Eds. 2001. *Handbook of Action Research: Participative Inquiry and Practice*. Trowbridge, UK, SAGE Publications.

Schuler, D., and A. Namioka. 1993. *Participatory Design: Principles and Practices*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Sheldon, W. M., M. A. Moran, and J. T. Hollibaugh. 2002. Efforts to Link Ecological Metadata with Bacterial Gene Sequences at the Sapelo Island Microbial Observatory. Pages 402-407 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.

Silverman, D. 1993. *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction.* London, Sage Publications, Inc.

Silverman, D. 1997. *Qualitative Research: Theory, Method and Practice.* Sage Publications, New Delhi.

Smith, D. J., B. J. Benson, and D. F. Balsiger. 2002. Designing Web Database Applications for Ecological Research. Pages 408-413 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.

Spycher, G., J. Cushing, D. Henshaw, S. G. Stafford, and N. Nadkarni. 1996. Solving problems for validation federation and migration of ecological databases. Pages 695-700. *Proceedings of Eco-Informa Workshop, Global Networks for Environmental Information, 4-7 November '96, Lake Buena Vista, FL*. Environmental Research Institute of Michigan (ERIM), Ann Arbor, MI.

Stafford, S. G., J. W. Brunt, and W. K. Michener. 1994. Integration of scientific information management and environmental research. Pages 3-19 in W. K. Michener, J. W. Brunt, and S. G. Stafford, eds. *Environmental information management and analysis: ecosystem to global scales*.

Stafford, S. G., N. E. Kaplan, and C. W. Bennett. 2002. Through the Looking Glass: What do we see, What have we learned, What can we share? Information Management at the Shortgrass Steppe Long Term Ecological Research Site. Pages 413-419 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics, Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.

Star, L. 1991. The sociology of the invisible: the primacy of work in the writings of Anselm Strauss. Pages 265-284 in D. R. Maines, ed. *Social organization and social process: essays in honor of Anselm Strauss*. Aldine De Gruyter, New York.

Star, S. L. 1999. The Ethnography of Infrastructure. *American Behavioral Scientist* 43(3): 377-391.

Star, L., and K. Ruhleder. 1994. Steps toward an Ecology of Infrastructure: Complex problems in design and access for large-scale collaborative systems. *CSCW94: Transcending Boundaries: Proceedings of the Conference on Computer Supported Cooperative Work, 22-26 October*: 253-264.

Star, S. L., and G. C. Bowker. 2002. How to Infrastructure. Pages 151-162 in L. A. Lievrouw and S. Livingstone, eds. *The Handbook of New Media*. SAGE Publications, London.

Star, S. L., and A. Strauss. 1999. Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work, Computer Supported Cooperative Work. *The Journal of Collaborative Computing* 8: 9-30.

Stonebraker, M. 1994. Sequoia 2000: a reflection on the first three years. *IEEE Internet Computing*: 108-116.

Strauss, A., S. Fagerhaugh, B. Suczek, and C. Wiener. 1985. *Social organization of medical work*. The University of Chicago Press, Chicago.

Strebel, D. E., D. R. Landis, K. F. Huemmrich, J. A. Newcomer, and B. W. Meeson. 1998 The FIFE data publication experiment Pages 1277-1283. *Journal of the Atmospheric Sciences*

Strebel, D. E., B. W. Meeson, and A. K. Nelson. 1994 Scientific information systems: a conceptual framework Pages 59-85 in W. K. Michener, J. W. Brunt, and S. G. Stafford, eds. *Environmental Information Management and Analysis: Ecosystem to Global Scales* Taylor & Francis London

Suchman, L. 2000. Organizing Alignment: A Case of Bridge-building. *Organization* 17: 311-327.

Swanson, F. J., and R. E. Sparks. 1990. Long-Term Ecological Research and the Invisible Place. *BioScience* 40:

502-508.
Thorley, M. R., and P. N. Trathan. 1994. The history of the BIOMASS data centre and lessons learned during its lifetime. Pages 313-322 in S. Z. El-Sayed, ed. *Southern Ocean ecology: the BIOMASS perspective*. Cambridge University Press, New York.
Van de Castle, J., D. Pennington, T. Fountain, and C. Pancake. 2002. A spatial data workbench for data mining, analyses, and synthesis. Pages 420-424 in N. Callaos, J. Porter, and N. Rishe, eds. *Proceedings of the 6th World Multi-Conference on Systematics Cybernetics and Informatics, 14-18 July 2002, Orlando, FL*. IIIS, Orlando, FL.
Veen, C., C. A. Federer, D. Buso, and T. Siccama. 1994. Structure and function of the Hubbard Brook Data Management System. *Bulletin of the Ecological Society of America* 75: 45-48.
Wenger, E. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.
Worster, D. 1977. *Nature's Economy*. Cambridge University Press.

8 Appendices

8.1 Appendix. List of Acronyms
8.2 Appendix. Expectations of Data Managers, 1989 Data Managers' Report.
8.3 Appendix. LTER IM Workshop/Symposium Collections
EcoInforma 1996: Global Networks for Environmental Information
DIMES 1998: Data and Information Management in the Ecological Sciences
IIS/SCI 2002: Systemics, Cybernetics and Informatics

8.4 Appendix. LTER Network News Article - Fall 2001
8.5 Appendix. Public Project Web Page - 2002
8.6 Appendix. Introduction to Field Work Handout - February 2002
8.7 Appendix. Ethnographic Fieldwork Timeline - 2002
8.8 Appendix. DGO Conference Project Paper - May 2002
8.9 Appendix. LTER Information Manager Meeting Presentation - August 2002
8.10 Appendix. NSF BDEI Meeting Presentation - February 2003
8.11 Appendix. Databits Whirlwind News Article - Spring 2003
8.12 Appendix. TOL-LTER Program Description - Spring 2003
8.13 Appendix. ECSCW-CSSC Collaboration Workshop Preface - Summer 2003
8.14 Appendix. LTER Information Manager Meeting Presentation and Handout – September 2003
8.15 Appendix. LTER All Scientists Meeting Poster - September 2003

8.16 Appendix. Sociotechnical Good Reads (1999-2003)
8.17 Appendix. Information Management and Information Systems:
Conferences, Journals, Proceedings, & Bulletins
8.18 Appendix. Project Publications

8.1 Appendix: Acronym List

ASCII	American Standard Code for Information Interchange
BDEI	Biodiversity and Ecosystem Informatics NSF Program
CC	Coordinating Committee (of LTER Site Representative Principle Investigators)
ClimDB	Climate Database
CoP	Community of Practice
EML	Ecological Metadata Language
ESA	Ecological Society of America
GS	Graduate Student
HSD	Human Social Dynamics NSF Program
IBP	International Biological Program
ILTER	International LTER
IM	Information Management
IMC	IM Committee
IMer	Information Manager
IMExec	IM Executive Committee
IT	Information Technology
LNO	LTER Network Office
LTER	Long-Term Ecological Research
MSI	Minimum Standard Installation
NIS	Network Information System
NISAC	Network Information System Advisory Committee (Isn't it NISAC?)
NSF	National Science Foundation
OBFS	Organization of Biological Field Stations
PD	Participatory Design
PI	Principle Investigator
QA/QC	Quality Assurance, Quality Control
RA	Research Assistent
SSC	Science Steering Committee
SLTER	Schoolyard LTER
T	Technician
USGS	United States Geological Survey

8.2 Appendix: 1989 Data Managers' Report (Excerpt)

http://intranet.lternet.edu/archives/documents/reports/committee_reports/Data-management-committee/1989-DM-committee-report/im_89_report.htm

EXECUTIVE SUMMARY

The August 1989 LTER Data Managers' workshop focused on developing solutions and avenues of communication to solve commonly experienced problems and meet future technological challenges. The meeting succeeded in creating a broad consensus about what the major issues were and how solutions were to be pursued. An agenda (Appendix A) was employed which maximized group interaction and discussion. The Data Managers from all seventeen LTER sites attended and actively participated (Appendix B).

…

(Page5)

Expectations of Data Managers

A memo was sent from Jerry Franklin (NET) to all Principal Investigators of record directing them to discuss their expectations of data management with their data manager prior to this meeting. Many did, however, not all.

Participants were asked to synthesize those discussions and assess them with regard to their own expectations and their perception of the expectations of the NSF. Four working groups were formed with group leaders assigned the responsibility to report back to the main group-as-a-whole with the consensus of the subgroup. Out of these discussions there were three recommendations extended for consideration by the LTER/CC.

1) Data management be recommended for elevation to a core area of LTER. This in light of the critical importance of data management to the overall goals of long-term ecological work and the increasing importance data management plays to intersite work. As more and more data are incorporated into databases there will be increasing demands placed on data management staff. These increases are already being felt at older sites.

2) Discussion and synthesis of the criteria by which data management should be evaluated in the NSF review process be included as an agenda item for the 1990 Data Managers' Workshop. PI's expect data managers to identify and act on these review criteria. The Data Managers should be in a position to make recommendations to the LTER/CC directed at establishing these criteria.

3) A member of the Data Management Task Force be included on the LTER/CC to provide a liaison to the group whose charge would be to provide realistic expectations of data management and represent the collective group of Data Managers. The Data Managers could better address issues of importance to network.

8.3 Appendix: LTER Information Management Workshop/Symposium Collections

Data and Information Management in the Ecological Sciences: A Resource Guide, 1998. Albuquerque, NM. W.Michener, J.Porter, and S.Stafford (eds). LTER Network Office, University of New Mexico. http://lternet.edu/documents/data-informationmanagement/DIMES/html/frame.htm

1. ISSUES AND CONCEPTS OF DATA MANAGEMENT - Susan G. Stafford

2. HARDWARE - Scott E. Chapal

3. TECHNOLOGICAL UNDERPINNINGS: COMMUNICATIONS AND NETWORKING - Rudolf Nottrott

4. TECHNOLOGICAL UNDERPINNINGS: SOFTWARE - Karen S. Baker

5. DATA ENTRY - John M. Briggs

6. DATA QUALITY CONTROL / QUALITY ASSURANCE - Don Edwards

7. SCIENTIFIC DATABASES FOR ENVIRONMENTAL RESEARCH - John H. Porter

8. ECOLOGICAL METADATA - William K. Michener

9. DATA ARCHIVAL - Richard J. Olson and Raymond A. McCord

10. THE WORLD WIDE WEB AS A TOOL FOR ECOLOGICAL RESEARCH PROGRAMS - Barbara J. Benson

11. PROVIDING INFORMATION ON THE WORLD WIDE WEB - John H. Porter

12. WEB-BASED DATA MANAGEMENT - Matthew B. Jones

13. VIRTUAL WORKING GROUPS AT NCEAS: USING THE WEB TO FACILITATE SCIENTIFIC COLLABORATION - Mark P. Shildhauer

14. VISUALIZATION OF ECOLOGICAL AND ENVIRONMENTAL DATA - John J. Helly

15. INFORMATION ACCESS AND DATABASE INTEGRITY AT THE NORTH TEMPERATE LAKES LONG-TERM ECOLOGICAL RESEARCH PROJECT - Barbara J. Benson and Maryan Stubbs

16. EVOLUTION OF THE KONZA PRAIRIE LTER INFORMATION MANAGEMENT SYSTEM - John M. Briggs

17. PALMER LTER INFORMATION MANAGEMENT - Karen S. Baker

18. MANAGEMENT OF A LONG-TERM WATER QUALITY DATABASE: FLATDAT FOR THE FLATHEAD LAKE BIOLOGICAL STATION - Melissa E. Holmes and Geoffrey C. Poole

19. THE H. J. ANDREWS CLIMATOLOGICAL FIELD MEASUREMENT PROGRAM - Donald L. Henshaw, Frederick A. Bierlmaier and Hazel E. Hammond

20. CLIMATE DATABASE PROJECT: A STRATEGY FOR IMPROVING INFORMATION ACCESS ACROSS RESEARCH SITES - Donald L. Henshaw, Maryan Stubbs, Barbara J. Benson, Karen S. Baker, Darrell Blodgett and John H. Porter

21. DATA AND INFORMATION MANAGEMENT IN THE ECOLOGICAL SCIENCES: SYNOPSIS FROM A FIELD STATION PERSPECTIVE - Hilary M. Swain and William K. Michener

Proceedings of Eco-Informa Workshop, Global Networks for Environmental Information, 4-7 November 1996, Lake Buena Vista, FL. Ann Arbor, MI: Environmental Research Institute of Michigan (ERIM)

Baker KS. 1996. Development of Palmer Long-Term Ecological Research Information Management. Pages 725-730.

Benson B. 1996. The North Temperate Lakes Research Information Management System. Pages 719-724.

Porter J, Hayden B, Richardson D. 1996. Data and information management at the Virginia Coast Reserve Long-Term Ecological Research Site. Pages 731-736.

Porter, JH., RW.Nottrott, and K.Baker, 1996. Tools for Managing Ecological Data. Pages 87-92.

Brunt JW, Nottrott R. 1996. The LTER network information system for the 21st century. Pages 104-104.

Olson R, Voorhees L, Field J, Gentry M. 1996. Packaging and distributing ecological data from multisite studies. Pages 93-102.

Stafford SG, Brunt J, Benson BJ. 1996. Training environmental information managers of the future. Pages 111-122.

Spycher G, Cushing J, Henshaw D, Stafford SG, Nadkarni N. 1996. Solving problems for validation federation and migration of ecological databases. Pages 695-700.
Stubbs M, Benson B. 1996. Query access to relational databases via the World Wide Web. Pages 105-109.

Wasser, C. Dynamic Data Transfer Via the World Wide Web: Increasing your Visitors' Understanding of Ecological Data. Pages 737-742.

Proceedings of The 6th World Multiconference on Systematics, Cybernetics, and Informatics July 14 - 18 2002 Orlando Fl (http://www.iiis.org/sci2002)

LTER session title: The Ecoinformatics Challenge: Meeting Ecological Information Needs for the Site, Network, and Community (LTER session co-chairs: John Porter, Karen Baker, Susan Stafford)

The future of ecoinformatics in Long Term Ecological Research
James W. Brunt, Peter McCartney, Karen Baker, and Susan G. Stafford
Emerging information technologies allow new exploration into tools for the management and use of information that solve problems for ecologists and create new and innovative lines of scientific inquiry. Collaborative, multi-disciplinary research programs to facilitate these new lines of inquiry have produced a need for scientific information systems that communicate data, information, and knowledge across spatial, disciplinary, and cultural boundaries.

Designing Web Database Applications for Ecological Research
Dan J. Smith, Barbara J. Benson, and David F. Balsiger
Many sites conducting ecological research must routinely manage a diverse suite of datasets and make them accessible to researchers. This paper presents an approach to creating an ecological data query system that dynamically creates predefined dataset query interfaces for managed datasets. The query interfaces are created from stored dataset metadata and query creation metadata and include only those field selections and filtering options identified as relevant for the specified dataset.

Organizational Informatics: Site Description Directories for Research Networks
Karen S. Baker, James W. Brunt and David Blankman
A site description directory plays a central role as a catalog for a network of research sites. Such a directory represents a keystone element in an information management system. A directory contributes to community communications both through documentation of member information and relationships as well as through design feedback elicited from participants in the ongoing process of developing the catalog system.

Through the Looking Glass: What do we see, What have we learned, What can we share? Information Management at the Shortgrass Steppe Long Term Ecological Research Site
Susan G. Stafford, Nicole E. Kaplan, and Christopher W. Bennett
This paper documents the development of a successful information management system at a Long Term Ecological Research (LTER) site that has a rich history of data collection and management. landscape. Early research projects focused on understanding native plants, recovery on abandoned plowed fields, and techniques for measuring plants' responses to grazing by cattle.

Efforts to Link Ecological Metadata with Bacterial Gene Sequences at the Sapelo Island Microbial Observatory
Wade M. Sheldon, Mary Ann Moran and James T. Hollibaugh
The existence of public databases for archiving genetic sequence data, such as GenBank and the Ribosomal Database Project, coupled with the availability of standardized sequence alignment and comparison tools has led to rapid advances in the field of bacterial genetics and systematics.

Ecological Informatics: a Long-Term Ecological Research Perspective
William K. Michener, James W. Brunt and Kristin L. Vanderbilt
Scientists within the Long-Term Ecological Research (LTER) Network have provided leadership in ecological informatics since the inception of LTER in 1980.

Common Information Management Framework: in Practice
Eda C. Meléndez-Colom and Karen S. Baker
A common goal of information management systems (IMS) is to share information among its users and originators. These systems are usually implemented by project managers and sponsors.

Transition from a Legacy Databank to an Integrated Ecological Information System
Donald L. Henshaw, Gody Spycher, and Suzanne M. Remillard
Many tasks and issues are encountered in the process of converting a scientific databank containing multiple legacy and long-term study databases into an integrated data production and distribution system.

Using XML-encoded Metadata as a Basis for Advanced Information Systems for Ecological Research
Peter H. McCartney and Matthew B. Jones
Metadata provide information on the structure and meaning of data. It is one of the most basic components for building a scalable, networked infrastructure for accessing ecological data. Several partnering groups from ecology have collaborated to define a standardized format for metadata that is machine-parseable and extensible.

A Spatial Data Workbench for Data Mining, Analyses, and Synthesis
John Vande Castle, Deana Pennington, Tony Fountain and Cherri Pancake
Information managers at ecological research sites grapple with the complexity of diverse and heterogeneous datasets. The effective management of large geospatial datasets requires extensive hardware, software, and human resources that are often beyond the capabilities of smaller institutions.

Integrating Ecological Data: Tools and Techniques
John H. Porter and Kenneth W. Ramsey, Jr.
Integration of data is critical to achieving new levels of understanding of ecological systems and processes. Typically, data integration is achieved only through a painstaking manual process that rules out large-scale integration.

Template-driven End-User Ecological Database Design
Judith Bayard Cushing, Nalini Nadkarni Keri Healy, Erik Ordway, Lois Delcambre, and Dave Maier
Historically, ecologists have collected and stored data in individualist ways, making data sharing among collaborators and subsequent data mining difficult.

8.4 Appendix: LTER Network Newsletter Fall 2001: Incubation of a New Idea

Designing an Infrastructure for Heterogeneity of Ecosystem Data, Collaborators, and Organizations

Karen Baker (UCSD, Palmer LTER) and Geoffery Bowker (UCSD, Communications Dept)

As more scientists exchange and post data digitally, the challenges of transforming data systems into knowledge systems come into sharper focus. As communities address issues of data availability, the understanding of data incompatibility issues deepens. The LTER integrative approach to science brings with it a familiarity with difficulties arising from technology with differences in hardware, software, and formats as well as from uniformity or standards with differences in collection units, classification categories, and methods.

There are a variety of challenges to data integration arising from creation of appropriate metadata, the descriptive documentation about the data. The 'context' of data is complex: one aspect of a dataset's metadata relates to the environment it describes while another aspect relates to its creator and is impacted by its association with a site, project, repository and network. As a result, in addition to the barriers of technology and uniformity, we identify within the multiple layers of metadata that there are barriers of social organization to consider.

A recent NSF report (Kinzig et al, 2001) focusing on priorities for interdisciplinary environmental research presents the human interface with the environment representing humans as agents within the environment as well as in its perception. The report calls for careful consideration of how scientists negotiate their data. How the environment or the data is reported with its attendant metadata impacts ultimately how it is perceived and reused. The realization that metadata are not socially or culturally neutral improves our ability to design knowledge systems. Techniques used in the structuring of information include classification schemas as well as attention to semantic vocabularies and domain maps to describe the context of the data.

We ask whether there are methods complementary to logic-based approaches to information retrieval that can include data collection practices, project summaries and research vignettes.

We were recently awarded a Biodiversity and Ecosystem Informatics (BDEI) cross-agency (NSF/USGS/NASA) (Maier et al, 2001) incubation grant to consider impacts on metadata and infrastructure of the heterogeneity of 1) data, 2) collaborators, and; 3) organizations. This project seeks new ways of grounding environmental data with metadata so that data can be used more flexibly today and also be available over the long-term in a form useable for unanticipated future queries. The goal of this project is twofold: to open up a new field of database enquiry tied to the specific challenges of biodiversity and ecosystem science and to initiate a cross-domain dialogue between LTER ecologists (Callahan 1984), information managers (Baker et al. 2000), and social scientists (Bowker 2001).

This BDEI incubation grant will support a postdoctoral fellow trained in participatory design techniques who will focus on a prototype project as well as a community forum. Work begins in January 2002 with arrival of Dr. Helena Karasti (2001) who recently finished her thesis work in the Department of Information Processing Science at Oulu University in Finland. With experience in collaborative design, she will work with selected data from the Palmer LTER while considering the data capture and metadata roles in transforming data into knowledge. This project will develop into a larger follow-on study of modes of packaging data. This work facilitates a timely dialogue focused on "data ecology" (the relationship between data and their multiple environments) and builds toward the concept of an "organizational ecology" (the relationship between data and participants and their networks). The potential impacts are high in terms of both network infrastructure and expert systems.

The LTER Network fosters an environment that engages research scientists in the process of preserving their data digitally for the long-term. The vision broadens and deepens with collaborative participation in metadata and knowledge system design. An active partnership spanning several communities (NCEAS/SDSC/LTER) currently is exploring a metadata-based framework and includes participation by the LTER Central Arizona-Phoenix site exploring structure and developing tools. Our BDEI project joins these informatics efforts. Given the many facets of human knowledge, it will take a team bridging data-creators and data-managers, working in collaboration with data-handlers as well as those familiar with data-perception and data-context to address data integration in the short-term and data reuse in the future.

The LTER Network is well suited to initiate studies on the use of technology and standards as well as on the impact of social organization on knowledge systems.

References

Baker, KS, BJ Benson, DL Henshaw, D Blodgett, JH Porter, and SG Stafford, 2000. Evolution of a multisite network information system: the LTER information management paradigm. BioScience 50(11): 963-978.

Bowker, GC, 2001. Biodiversity, Datadiversity. Social Studies of Science, 30(5), 643-684.

Callahan, JT, 1984. Long-Term Ecological Research. BioScience 34(6), 363-367.

Karasti, H, 2000. Increasing sensitivity towards everyday work practice in system design. Unpublished PhD thesis. http://www.tol.oulu.fi/~helena/

Kinzig, AP, SCarpenter, MDove, GHeal, SLevin, JLubchenco, SHSchneider, DStarrett, 2000. Nature and society: an imperative for integrated environmental research Excecutive Summary of a workshop sponsored by NSF, Developing a Research Agenda for Linking Biogeophysical and Socioeconomic Systems, Tempe, AZ, 5-8 June, http://lsweb.la.asu.edu/akinzig/report.htm

Maier, D, E.Landis, JCushing, A.Frondorf, ASIberscatz, MFrame, and JLSchnase, 2001. Research Directions in Biodiversity and Ecosystem Informatics, pp30. Workshop held at NASA Goddard Space Flight Center, June 22-23, 2000, Greenbelt, Maryland.

8.5 Appendix: Public Project Web Page

NSF Digital Government:
Biodiversity and Ecosystem Informatics 2002

http://interoperability.ucsd.edu/project/02dgo/

Project Title:
Designing an Infrastructure for Heterogeneity of Ecosystem Data, Collaborators and Organizations

BDEI Team
Geof Bowker (Publication Data Diversity)
Karen Baker (Publication LTER IM)
Helena Karasti: (Publications PhD Thesis 2001; Teleradiology 1998)

Activities
2001 BDEI Proposal: Summary
01Sep LTER Network News Fall01: Article
02Feb: LTER PI Introduction Fieldwork Handout
02May DGO: Conference (Announcement): Paper(PDF); Slides(PPT)
02Jun High Tech Forum (Finland): Article
02Jun NSF Nuggets: Slides(PPT)
02Nov CSCW: Conference; Position Paper Draft
03Jan ACM SIGGROUP: Bulletin- Paper, Volume
03Feb BDEI: Workshop, Summary-1page, PPT
03Apr LTER IM Newsletter DataBits: Whirlwind Tour (txt), GoodReads (txt)
03Apr US/Finland: Videoconference Initiation
03Apr TOL-LTER Graduate Students
03Jun UCSD/SIO: Time Capsule and Long-Term Data
03Sep Finland: High Tech Forum News Article; Bowker Lecture; Oulu Daily News Article
03Sep ECSCW Conference: CSSC Workshop Announcement; (W10 Workshop); Agenda (pdf), Proceedings Preface, Proceedings Report
03Sep CSSC SIGGROUP Paper
03Sep LTER IM Newsletter DataBits: GoodReads (txt)
03Sep LTER ASM: IM Abstract, Poster; Presentation/Handout Page1 & Page2
03Sep LTER Information Manager Meeting Presentation (ppt)
04Jan HICSS : Abstract , Paper(pdf)
04Feb eScience Requirements Capture Workshop paper
05Apr JIIS Paper (pdf) For continuation of this work, see Comparative Interoperability Project.

Ethnographic Field Work
Jan02: LTER Palmer fieldwork (ongoing)
Feb02: LTER Palmer PI Meetings (San Diego, ongoing)
Feb02: LTER IM Executive Committee Meeting (San Diego, SDSC)
Apr02: LTER Site Visit - SEV
Apr02: LTER CC Executive Committee Meeting (Sevilleta)
Apr02: LTER CC Committee Meeting (Sevilleta)
Apr02: LTER Network Office Visit - NET (Albuquerque)
Jun02: LTER Site Visit - CAP
Jun02: LTER IM Metadata Workshop (Phoenix)
Jul02: LTER IM Committee Meeting (Orlando)
Aug02: LTER AND Site Visit (Oregon)

Sep02: LTER PAL PI Meeting (Virginia)

Other
Digital Government Organization DG.O
BDEI00 Workshop
LTER Network
LTER Palmer

Support:
NSF BDEI 2002: Biodiversity and Ecosystem Informatics
in collaboration with
NSF OPP: Palmer Long-Term Ecological Research Program
NSF DBI: Palmer LTER Site Information Management: Interfaces for Science and Education Interoperability

Contact: web manager

8.6 Appendix: Introduction to Fieldwork - February 2002

Managing Heterogeneity: Ethnographic Fieldwork within the LTER

Geoffrey Bowker, Karen Baker and Helena Karasti

"Where is that data file?" "How come no one told me about a meeting?" These are just two examples of the everyday questions we face as information flows across our desk. A 2002 study will focus on the management of heterogeneous data, collaborators and organizations as we explore how long-term ecological research attends to the handling and the preservation of both scientific and organizational information. LTER researchers as stakeholders and members of a collaborative network have unique experience with and insight into existing and emerging practices that are pertinent to the design of future research environments. As integrative science inquiries broaden and cross domains, the practice of science often involves participation in a range of collaborations through individual, project, discipline and digital networks.

With the support of a one year NSF/BDEI grant at UCSD (Designing an Infrastructure for Heterogeneity of Ecosystem Data, Collaborators and Organizations – G.Bowker and K.Baker), postdoctoral researcher Helena Karasti will gather an understanding of the existing heterogeneous "data ecologies" and "data practices" of LTER. In her recent thesis Karasti studied work practices and systems development in the field of clinical radiology, and explored ways of bridging the gap between the worlds of technology use and design. In the LTER context, her work will include promotion of community discussion and exploration of those aspects of organizational information where preservation of context alongside site data enhances scientific legacy. Fieldwork will begin with interviews and participant observation with members of Palmer LTER and the LTER network. Though the starting point is with individual people and their work, the research interest is not in practitioners' personalities or individual styles but rather in their mundane, everyday information practices, e.g. how work is carried out, how information is generated and stored, how collaboration is achieved, and how technical and organizational infrastructure is used. The success of larger-scale collaborative efforts depends upon the production of high quality science through working practices facilitated by a supportive infrastructure. The fields of computer supported cooperative work (CSCW) and social informatics provide tools for developing a useful understanding of how we do long-term and networked science. The "Managing Heterogeneity" project will focus on the ways in which both scientific and organizational data is produced, used, shared, stored and reused across a range of media (such as paper, documents, journals, email, computer networking, organizational memory) within the LTER.

Helena Karasti can be contacted by email: hkarasti@ucsd.edu, work phone: (858) 822 5791

*References on the importance of work practice and user participation with design:
Baker and Bowker, 2001: http://intranet.lternet.edu/archives/documents/Newsletters/NetworkNews/fall01/fall01_pg09.html
Bowker and Star, 2001: Sorting Things Out, Classification and Its Consequences
Karasti et al, 1998: http://pal.lternet.edu/biblio/finalms/karasti_etal98_teleradiology.pdf
Karasti, 2001: http://pal.lternet.edu/biblio/finalms/karasti_helena_thesis2001.pdf

8.7 Appendix: Ethnographic Fieldwork Timeline 2002

Ethnographic Fieldwork Timeline 2002

02BDEI- 01-31958 NSF DGO

2002: Human Subject Forms

2002: 02BDEI Meetings (Jan 02, May 02, Jun 02, Sep 02)

Jan 02: LTER PAL Site Visit (ongoing)

Feb 02: LTER Palmer PI Workshop (San Diego, ongoing)

Feb 02: LTER IM Executive Committee Meeting (San Diego, SDSC)

Apr 02: LTER SEV Site Visit

Apr 02: LTER CC Executive Committee Meeting (Sevilleta, New Mexico)

Apr 02: LTER CC Committee Meeting (Sevilleta, New Mexico)

Apr 02: LTER Network Office Visit - NET (Albuquerque, New Mexico)

Jun 02: LTER CAP Site Visit - (Phoenix, Arizona)

Jun 02: LTER IM Metadata Workshop (CAP, Phoenix)

Jul 02: LTER IM Committee Meeting (SCI2002, Orlando)

Aug 02: LTER AND Site Visit (Corvallis, Oregon)

Sep 02: LTER PAL PI Meeting (Duck, North Carolina)

8.8 Appendix: DGO BDEI Project Paper

Designing an Infrastructure for Heterogeneity in Ecosystem Data, Collaborators and Organizations

Proceedings of the Second National Conference on Digital Government Research, 141-144pp, 20-22 May20002, Los Angeles

Karen S. Baker

Scripps Institution of Oceanography
University of California, San Diego
La Jolla, California 92093-0218 USA
mailto:kbaker@ucsd.edu

Geoffrey C. Bowker and Helena Karasti

Department of Communication
University of California, San Diego
La Jolla, California 92093-0503 USA
bowker@ucsd.edu, hkarasti@ucsd.edu

1. Abstract

To develop robust datasets for long-term re-use, new approaches are needed that incorporate relevant facets of organizational culture in their description. Early ethnographic research points to the importance of holding narrative accounts of data use alongside formal metadata structures. We describe our proposal to identify models for the design of information protocols and procedures within the Long-Term Ecological Research community that take account of the working practices of all the participants involved in the varied aspects of information processing.*

2. Introduction

Biodiversity and ecosystems data are currently being gathered in a large range of formats by a constellation of loosely connected private, government and not-for-profit agencies. The normal response to this double heterogeneity has been the development and enforcement of metadata (data about data) standards; in this response one tries to abstract data away from its organizational context in order to render it universally accessible. This project takes the opposite tack, and seeks new ways of grounding environmental data in its organizational context in such a way that it can both be used more flexibly today and so it can retain value longer. The hypothesis, based on the last 25 years of work in the field of Science Studies, is that formal data descriptions must be 'wrapped' in informal descriptions in order to be useful. The goal of this project is to open up the database inquiry of the biodiversity and ecosystems communities generated by their need for very long lasting and highly distributed data. We focus on communications in ecosystem informatics through the use of structural (e.g. standardized classifications; metadata) and alternative (e.g. narrative) methods. Our approach is action-oriented research that integrates ethnographic fieldwork and participatory design (Karasti, 2001). Through our theoretical interests in information ecologies and work practices, we intend to articulate connections between organizational and scientific data.

3. Research Approach

* NSF Grants EIA-01-31958, DBI-01-11544 and OPP-96-32763 support this work.

The issues involved in biodiversity and ecoinformatics are complex and large-scale. A recent call for setting priorities for new interdisciplinary environmental research programs points out the need for action outside the status quo of disciplinary science (Kinzig et al, 2000). We have gathered a team of investigators that shares the recognition that contemporary research questions require new interdisciplinary approaches.

Odum (1996) writes "because ecology is an integrative science, it has tremendous potential to provide a communication bridge between science and society". We extend his observation to include the bridging between ecological field sciences, information sciences and social sciences. We are working at the intersection of these three sciences taking into account the distinct Communities of Practice (CoP, Lave and Wenger, 1990) that have developed at the intersections (Figure 1). Active boundary communities include those interfacing Environmental and Social Sciences (e.g. Computer Mediated Communication, CMC), Environmental and Information Sciences (e.g. Information Management, IM), and Information and Social Sciences (e.g. Human Computer Interaction, HCI; Computer Supported Cooperative Work, CSCW; Social Informatics, SI; and Participatory Design, PD) with new insights into data handling, work practice and infrastructure effectiveness. The Center (U) designates a union of understanding.

3.1 At the Interface of Information Sciences and Social Sciences

The number of CoPs shown in Figure 1 at the join between Information Sciences and Social Sciences is an indicator of the complexity that is socially generated (PCAST, 1998). Social Informatics refers To the body of research and study that examines social aspects of computerization, including the roles of information technology in social and organizational change and the ways that the social organization of information technologies are influenced by social forces and practices (http://www.slis.indiana.edu/si/). Computer Supported Cooperative Work is a term to describe the understanding of the way people work in groups with the enabling technologies of computer networking, and associated hardware, software, services and techniques. It is concerned with designing shared information spaces and supporting heterogeneous, open information environments that integrate existing single-user applications (http://www.telekooperation.de/cscw/cscw.html). Participatory Design is an approach to the design and development of technological and organizational systems that places a premium on the active involvement of workplace practitioners in design and decision-making processes (http://www.cpsr.org/program/workplace/PD.html)

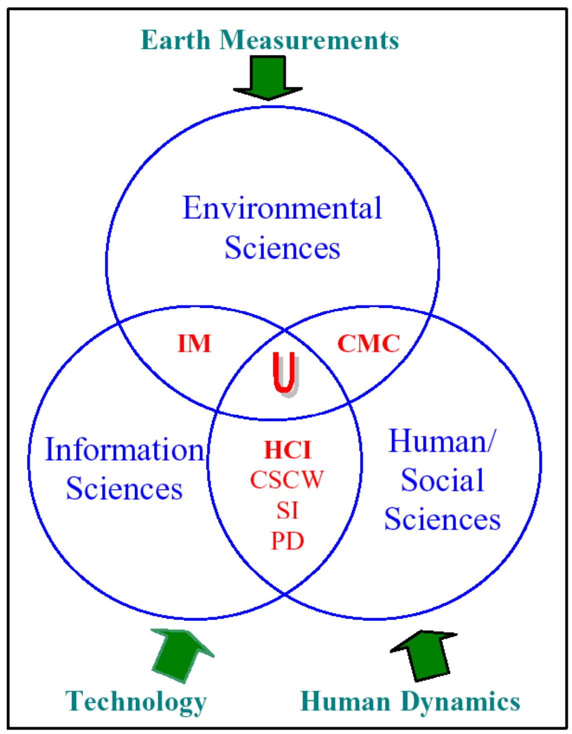


Figure 1: Science Domains and Communities of Practice (CoP)

Our work represents one approach to examining the issues of managing data for long-term, and eventually very long-term, use within a widely distributed, loosely connected network of scientists:

the Long-Term Ecological Research (LTER) network (Franklin, 1990). We have argued

(Bowker, 2001; Baker et al, 2000) that full attention must be given to the social and organizational dimensions of ecosystem informatics if we are to develop robust, reliable and useful databases for the future. Our research effort is grounded on ethnographic fieldwork. Ethnography is an approach for developing understandings of the everyday activities of particular communities of people through participant observation and interviews. The LTER is a complex working environment given the diversity of twenty-four sites, a Network Office coordinating the sites, and a range of associated partnerships. Embedded within each site and the Network Office is an active information management component. We are studying the data ecologies and work practices within the LTER, eliciting and articulating significant elements of collaboration and community, and considering design artifacts that enhance network science in order to better understand and to plan for the management of scientific heterogeneity.

How can the goal of creating and preserving meaningful long-term data be initiated by and grounded in everyday practice? We aim to articulate the relationships between data held in the computer and transmitted over a network, data held in the human mind and shared through stories as well as data held on paper and stored in file cabinets. A workshop format is planned with members of the LTER community in order to elicit multiple voices and to promote community discussion. We see workshops as an important enabling mechanism for reflection and change.

3.2 Research Focus

There is a wide gap in the field of ecosystem informatics between what is being produced by information technology specialists and what is actually useful to working scientists. The difficulties inherent in bridging this divide in computer science and field science partnerships are documented in retrospectives (e.g. Stonebraker, 1994) but are rarely addressed directly in practice. An interdisciplinary team working with an established, ongoing network provides the unique opportunity to focus on communications in ecosystem informatics.

The flow of information in field sciences consists of multiple steps starting with project design followed by fieldwork, moving data through ordering filters or frameworks into a digital record that may then be put through an output filter for retrieval (figure 4, Baker et al, 2000). In most scientific databases, organizational data falls away and is lost very quickly. Bowser (1986) for example, discusses problems with interpreting data predating the LTER site in Wisconsin. Measurements of lake water acidity would be different depending on whether they were taken in the laboratory on return from the field or in the field – loss of CO_2 in samples over a few hours changes the measurements. This information was nowhere mentioned in published reports, but fortunately Bowser and colleagues were able to locate a retired limnologist who remembered the procedure. The point here is that knowledge about the *practice* of old limnologists was needed. No-one at that time would have thought to retain this information about the data that they were collecting – everyone using the data at the time would know how lake water was collected. However, this vital information was lost over time.

Clearly, metadata standards alone will not solve this kind of problem. There is an initial awareness of this within the LTER information managers, as one of them stated: "We are finding now that the structured [metadata] is much more useful in terms of producing machine readable information but the narrative often times contains more information." In this study we are starting to explore today's work procedures while considering those that will be vital to scientists fifty or one hundred years from now. We are contemplating ways of preserving organizational data (defined as data about synthesis, work practice and institutional framework) without overburdening the already stretched resources and time of research projects and data management.

We view databases as communication tools for sharing data. There are two categories of sharing: the here-and-now of data collection in support of ongoing ecological research as well as the future of data re-use in answering different, as yet unasked, questions. One information manager cognizant of these two aspects articulated the following: “if people feed us back information about a dataset … we put it into the database, then other people can read what other people have said about that dataset… some of these people in the past have given us really comprehensive reviews of the data, it’s like wow, this should be part of the data, I did not know that. I’ve not had time to analyze it, so if someone takes the time to analyze it, especially an outside person, a PI might tend to do some corrections or what ever, but someone outside really sees it objectively: this does not match, this does not make sense… Another thing … the data manager knows a lot about what really are the good and bad aspects of the data. … because we have handled it, we know what works and what does not… That should be part of the metadata. Because ultimately if you don’t write those things down, they are going to get lost. … It’s stuff that is more valuable than a lot of this other descriptive information about a dataset. I mean in terms of a real quality ‘gut feeling’ of how good it is. You know, like a ‘subjective quality indicator’ of some sort.”

Recognizing the incremental change processes inherent to long-term datasets, one information manager describes: “When we now are moving our datasets into our new system, we need to go back and see the abstracts written 20 or 30 years ago … we are not asking the same questions anymore. We need to keep the old ones and start writing more descriptive information because the thinking changes… People’s thoughts on why it is being collected and should continue to be collected, change - different research questions are being asked.”

This project starts to explore new ways of grounding environmental data in its organizational context so that it can both be used more flexibly today and so it can retain its value longer. It will develop into a larger follow-on study of the articulation between metadata and narrative modes of data and possible ways of representing them. This work facilitates a timely dialogue focused on “data ecology” (the relationship between data and their multiple environments) and builds toward the concept of an “organizational ecology” (the relations between data, participants and their networks).

3.3 Initial Findings

While we are still in the data collection cycle of our project, some initial themes are emerging from interviews and observations of work practice. The formal work practices of LTER information managers relate to gathering, quality analysis and quality control, archiving and facilitating data exchange. In themselves, these comprise a demanding set of tasks. However, even more demanding are the inherent, continuing tensions between the formal and the informal work of creating and holding the organizational memory encompassing local datasets and the network information in general. This suggests the need to find ways to characterize and represent informal work that would enable long-term data use. Within the Library and Information Science community is a growing “awareness of the immense scope of the potential preservation crisis” (http://www.clir.org/pubs89/contents.html) with ‘incremental metadata considered a key to successful migration of data. Such incremental metadata will take many forms in the attempt to determine the appropriate mix of structured and narrative accounts of datasets.

4. Conclusion

We propose to identify pertinent types of contextual information relevant to data synthesis and promote community discussion to enhance representation of dynamic, multi-level aspects of ecological data. As is the case with action research, a dual level approach is planned: both the practical level as represented by our work of learning with and observing the LTER as well as the conceptual level which transcends scientific community. Such an approach sets the stage for asking whether there are methods complementary to logic-based approaches to information retrieval that can encompass contextual understanding, including both lived experiences and historical understandings.

Acknowledgement: We offer special thanks to the LTER community members for contributing their time and insight.

5. References

Baker, K.S., B.J. Benson, D.L. Henshaw, D. Blodgett, J.H. Porter, and S.G. Stafford, 2000. Evolution of a multisite network information system: the LTER information management paradigm, Bioscience, 50(11), 963-978.

Bowker, G.C, 2001. Biodiversity Datadiversity, Social Studies of Science, 30(5), 643-684.

Bowser, C., 1986. Historic Data Sets: Lessons from the past, lessons for the future, Research data management in the ecological sciences W. Michener (ed.) University of South Carolina Press, The Belle W. Baruch library in marine science 16: 155-179.

Franklin, J.F., C.S. Bledsoe, J.T. Callahan, 1990. Contributions of the long-term ecological research program - an expanded network of scientists, sites, and programs can provide crucial comparative analyses, Bioscience 40(7): 509-523.

Karasti, H., 2001. Increasing sensitivity towards everyday work practice in system design. PhD thesis, http://herkules.oulu.fi/isbn9514259556.

Kinzig, A.P., S. Carpenter, M. Dove, G. Heal, S. Levin, J. Lubchenco, S.H. Schneider, D. Starrett, 2000. Nature and Society: An Imperative for Integrated Environmental Research, Workshop Report, http://lsweb.la.asu.edu/akinzig/report.htm.

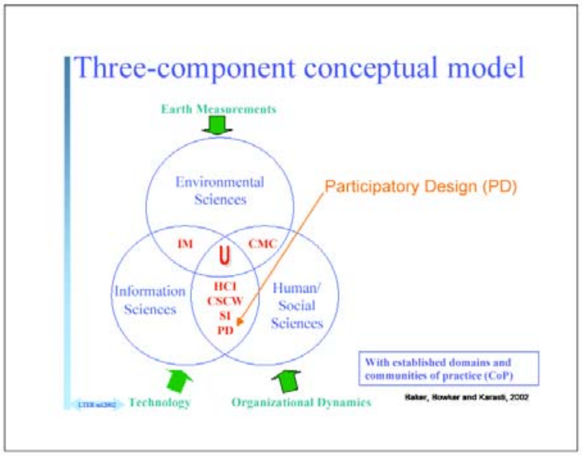
Lave, J. and E. Wenger (1990). Situated Learning: Legitimate Peripheral Participation. Palo Alto, CA, Institute for Research Learning: 1-38.

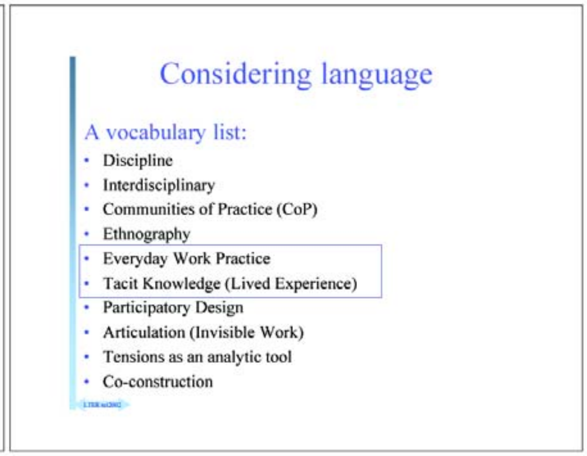
Odum, E., 1996. Ecology: A Bridge Between Science and Society. Sinauer Associates, MA.

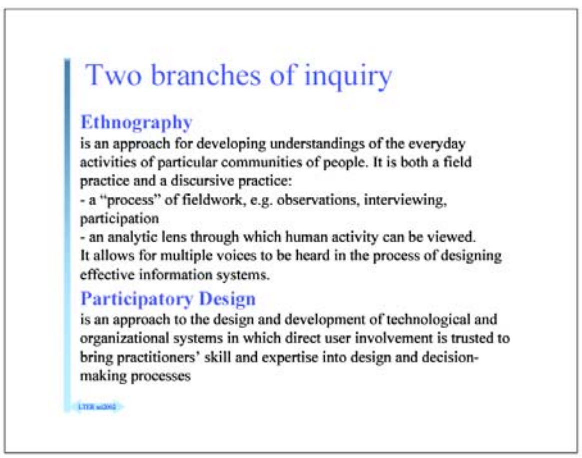
PCAST, 1998. Teaming with Life: Investing in Science to Understand and Use America's Living Capital. President's Comm of Advisors, Science and Technology, Panel on Biodiversity and Ecosystems.

Stonebraker, M., 1994. Sequoia 2000 -- A reflection on the first three years, Seventh International Working Conference on Scientific and Statistical Database Management, 28-30 September '94, Charlottesville, VA, pp 108-116. IEEE Computer Society Press, Los Alamitos.

8.9 Appendix: Information Manager Meeting Presentation – August 2002



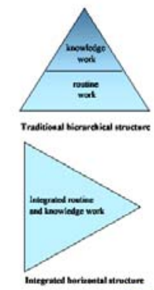




Seeing everyday practices

- not as "boring" routines
- but as routines that disclose the experienced and practised grasp of complex skills and competencies brought together to carry out ordinary work.

Organizational Models



Metadata in the LTER

"Metadata has become adapted as the terminology for almost everything except the scientific observation itself, everything that leads up to producing that number." (LTER IM 2002)

Contexts:

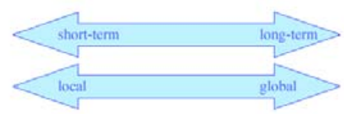
- EML, a technological solution
- scientific
- semantic, e.g. controlled vocabularies
- ontological
- organizational

It's important to articulate and maintain the multiple contexts!

Tension 1

"We are finding now that the structured [metadata] is much more useful in terms of producing machine readable information but the narrative often times contains more information." (LTER IM 2002)

Tensions 2



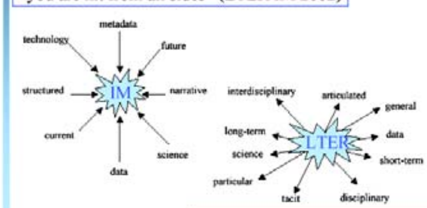
"The more it [data] is described with metadata, the more is known. Differently done if only for people at own site who know a lot about the collection system, etc. - but if done for someone outside or 20-30 years down the road, it gets more critical. In the near term you do not see the need to document obvious things." (LTER IM 2002)

Tension 3

"When we now are moving our datasets into our new system, we need to go back and see the abstracts written 20 or 30 years ago … we are not asking the same questions anymore. We need to keep the old ones and start writing more descriptive information because the thinking changes… People's thoughts on why it is being collected and should continue to be collected, change - different research questions are being asked." (LTER IM 2002)

Tensions and balances

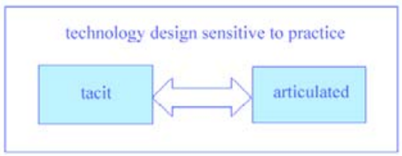
"you are hit from all sides" (LTER IM 2002)



"I feel like I am getting spread too thin" (LTER IM 2002)

Articulating the tacit for design

"There are many activities that are so much a part of our everyday lives that we are unable to provide accurate accounts even when asked to reflect upon these activities. In some cases, we may not have the vocabulary to talk about them." (Blomberg et al 1993)



IM lived experiences

"The data manager knows a lot about what really are the good and bad aspects of the data. … because we have handled it, we know what works and what does not… That should be part of the metadata. Because ultimately if you don't write those things down, they are going to get lost. … It's stuff that is more valuable than a lot of this other descriptive information about a dataset. I mean in terms of a real quality 'gut feeling' of how good it is. You know, like a 'subjective quality indicator' of some sort." (LTER IM 2002)

User experiences

"if people feed us back information about a dataset … we put it into the database, then other people can read what other people have said about that dataset… some of these people in the past have given us really comprehensive reviews of the data, it's like wow, this should be part of the data, I did not know that. I've not had time to analyze it, so if someone takes the time to analyze it, especially an outside person, a PI might tend to do some corrections or what ever, but someone outside really sees it objectively: this does not match, this does not make sense." (LTER IM 2002)

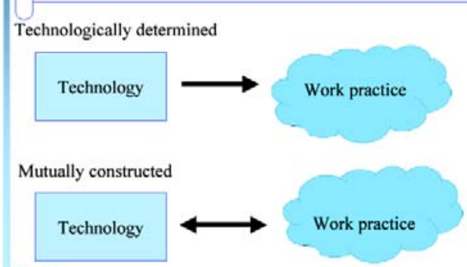
Articulation brings:

- understanding of heterogeneities
  - tensions & balances
  - knowledge & experience
- recognition of multiple IM roles
  - service
  - research partner
  - change agent

Metadata represents:

- a framework
- a developing structure

Invitation: to recognize these distinct processes



Invitation: to value and to articulate lived experiences and the everyday data/metadata practices

8.10 Appendix: NSF BDEI Meeting Presentation - February 10, 2003

An Emergent Global Biodiversity

Information Infrastructure: Social Informatics Research Agenda

G.Bowker, K.Baker and H.Karasti

NSF/BDEI 2002: Managing Heterogeneity: Data, Collaborators, Organizations

Biodiversity information is a key to the possible survival of our planet as we know it: we need to be able to aggregate data from multiple sources to produce a reasonable picture of where the planet is going, and in order to judge how effective different interventions may be. In fact, biodiversity informatics research furnishes a rich, distinctive arena for domain (in our case ecological field science), computer, and social scientists to work together.

Following our BDEI work, we see three key agendas for socio technical informatics research: negotiating infrastructure; evaluating infrastructural development; and exploring the fit between policy frameworks and information policies.

- *Negotiating Infrastructure:* The dream of a common infrastructure has accompanied scientific work for much of the past few centuries. A key contribution here for social analysis is to describe clearly the social, political, and ethical stakes which accompany information infrastructure.

- *Articulating Evaluation Procedures:* There are two interrelated sets of work involved in the development of large-scale biodiversity science infrastructures: **first** the development of a robust, flexible technological base and **second** changes in scientific culture. Infrastructure problems play out over twenty to thirty years while funding comes in three to five year chunks. The solution here is not to abandon the process of formative evaluation in order to speed things up but rather to evolve formative evaluation approaches in order to make key organizational and institutional interventions at the appropriate moment.

- *Policy Dimensions:* Moving toward the digital earth will be simultaneously fundamental for basic research and a very important policy tool. It is essential that attention be given to the layering of technical choices into models. Where values are incorporated, they best be tagged so a naïve end user will have an understanding of what is built in. This goes against one metric today where a successful interface hides complexities. A contribution here of the social scientific approach is to locate organizational and cultural innovations which have proven successful and to work with partnerships to scale these innovations.

We have developed cross-cutting concepts to analyze the relative and changing roles of organizational and technical decision making in developing cyberinfrastructure: protocols and metadata. ***Protocols*** provide the technical handshake between computers and the organizational arrangements between institutions. Due process analysis uncovers local workarounds and tacit knowledge. Protocols are an essential part of building large-scale cyberinfrastructures; they address the distribution between technical and due process fixes to a given problem. An expanded concept of ***metadata*** is a way to capture both technical aspects of a dataset and the institutional and organizational setting of the production of that dataset. Alternative forms of narrative are important for both data and organizational informatics. This information becomes ever more valuable as the dataset's developers move on, and as the information spreads across organizational contexts (from a scientific debate to a policymaker's input).

We address the biodiversity agenda working with the Long-Term Ecological Research network, an eco-socio-technical community. With data management embedded locally at each site, LTER is experiencing the digital infrastructure as a process. We use ethnographic and participatory design methods where our fieldwork included 52 interviews with information managers, scientists, technicians, administrators, managers and associates. We initiated dialogue with a presentation to the community on the role of information managers and Communities of Practice. An invitation was issued to consider the difference between technologically determined and mutually constructed approaches to IM. This is a timely invitation since the community is re-addressing the question of dataset documentation in the form of metadata. The ramifications of metadata decisions are broad ranging from existing classification schemes to developing applications to evolving semantic and ontological approaches.

So we come to the challenges for Biodiversity Informatics: to design for irreducible heterogeneity in domain, cultural, and policy arenas. Building bridges and conducting negotiations, infrastructure is an ongoing, collaborative work where we work to learn how to articulate new skills and new roles as well as how to gauge the progress of an emergent cyberinfrastructure along the multiple fronts it needs to succeed on – the cultural and organizational as well as the technical.

8.11 Appendix: Collaborative Practice, Databits Spring 2003

Databits - Spring 2003 (LTER Information Manager Newsletter)
http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/03spring/#1nb

A whirlwind tour of collaborative practice

Karen Baker and Helena Karasti, Palmer Station LTER

The Computer Supported Cooperative Work (CSCW) is a computer systems' research and development community that brings together the social and technical aspects for supporting collaboration. Karen Baker and Helena Karasti attended the biannual CSCW Conference in November 2002 (http://www.acm.org/cscw2002/). CSCW is sponsored by the Association of Computing Machinery (ACM). The ACM is an organization founded in 1947 (http://www.acm.org) to advance the skills of information technology (IT) professionals and students worldwide and it houses several special interest groups (SIGs) such as Groupware (SIGGROUP) and Management of Data (SIGMOD).

Karen and Helena participated at CSCW02 in a metadata-related workshop entitled "Storytelling and Collaborative Activities". A paper derived from our presentation will be published this year in the SIGGROUP Bulletin. In addition, we were able to attend several tutorials including "A Whirlwind Tour of CSCW Research", "Understanding Collaborative Activities and Applications: Methods for Studying Usefulness, Usability and Use of CSCW Systems", and "Collaboration Technology in Teams, Organizations, and Communities".

Helena has returned to the Oulu University in Finland, with her UCSD and LTER ties continuing. As a professor in the Department of Information Processing Science, she recently placed an ad (http://www.tol.oulu.fi/lter.html) for master's students to work with her on LTER materials collected during her year at UCSD. From the diverse ethnographic materials, final transcriptions of more than 50 interviews are being completed just this month.

In collaboration with Karen Baker and Geof Bowker, she has proposed to lead a workshop at the European Computer Supported Cooperative Work conference to be held in Helsinki, Finland this fall (http://ecscw2003.oulu.fi/). The workshop, titled "Computer Supported Scientific Collaboration", aims to bring together for the first time those interested in use of CSCW views and methods within the scientific arena. The workshop proposal is in recognition that much of the CSCW community work focuses on the business, medical, and education fields whereas the challenges that scientific collaborations pose for CSCW may be somewhat different.

The alignment of opportunities was most fortunate to support a one year study with the LTER community combining an ethnographic focus on technologically mediated work practices with participatory design (e.g. Information Systems Research in Scandinavia, http://iris.informatik.gu.se/). A report and paper are in preparation to serve as a continuation of the dialogue initiated with the LTER IM community in February 2002 at the IM Executive Committee meeting in San Diego and at the LTER IM Committee meeting in July in Orlando. Having introduced the concepts of sociotechnical systems and participatory design, plans are developing for Karen to visit Oulu in relation to attending ECSCW03 and for Helena to revisit San Diego. The visits allow continuation of analysis and co-writing, work with the students, and plans for future collaboration as well as of investigations into local floras, faunas and saunas.

8.12 Appendix: TOL-LTER Program Description - Spring 2003

Helena Karasti initiated work in April 2003 as thesis supervisor with five graduate students (http://www.tol.oulu.fi/lter.html) with the Department of Information Processing Science (TOL) at the University of Oulu. The group works in association with the Long-Term Ecological Research Program (LTER), continuing collaboration with Karen S. Baker and Geoffrey C. Bowker at University of California San Diego, Scripps Institution of Oceanography and Communications Department, respectively.

Areas of interest for the student theses include:

- Understandings of information technology in LTER
- Long-term databases and the development of metadata in LTER
- Long-term databases and data management: case of LIDET experiment
- Data sharing in LTER
- Information managers' career paths and expertise

8.13 Appendix: ECSCW03 –CSSC Report Summer 2003

European Computer Supported Scientific Collaboration Workshop Report Preface

Karasti, H., K.S.Baker, and G.C.Bowker (eds)
Proceedings of the Computer Supported Scientific Collaboration Workshop, Eighth European Conference on Computer Supported Cooperative Work
Helsinki, Finland, 14 September 2003
University of Oulu, Department of Information Processing Science
Research Paper Series A34, Oulu 2003

Preface

The Eighth European Computer Supported Cooperative Work conference (ECSCW 2003), a biannual forum that brings together the social and technical aspects for supporting collaborations, provides a venue this year to gather researchers interested in the study of scientific collaborations and their technology support.

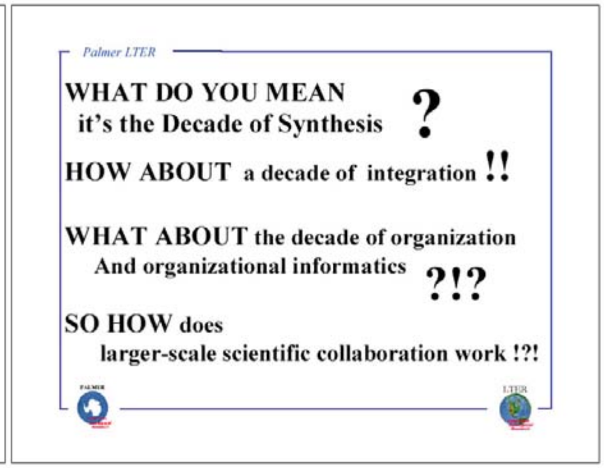
The response to the call for papers for the Computer Supported Scientific Collaboration workshop (CSSC) is evidence of the CSCW community members shared interest in the elements of collaboration particular to scientific communities and in the challenges they present for designing computer-based support systems.

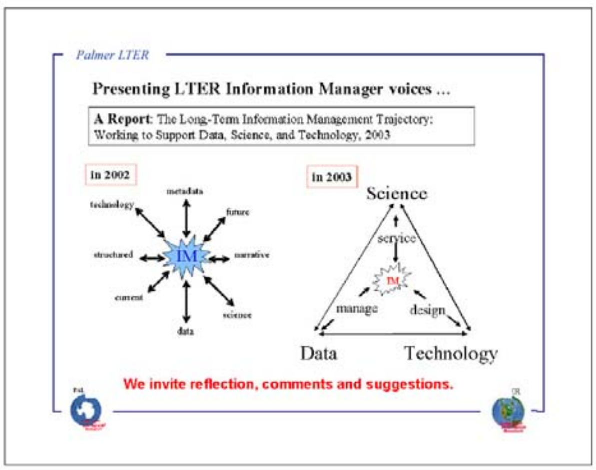
Like the Three Smiths of Nylund's statue in Helsinki, we three organizers came together to work jointly at crafting an understanding of a scientific network, the Long-Term Ecological Research program. Having hammered out a workshop agenda, we welcome the CSSC participants. Through the position papers collected here and with our diverse case studies, from ecological teams and human genes to digital streams, we add shape to the characterization of CSSC.

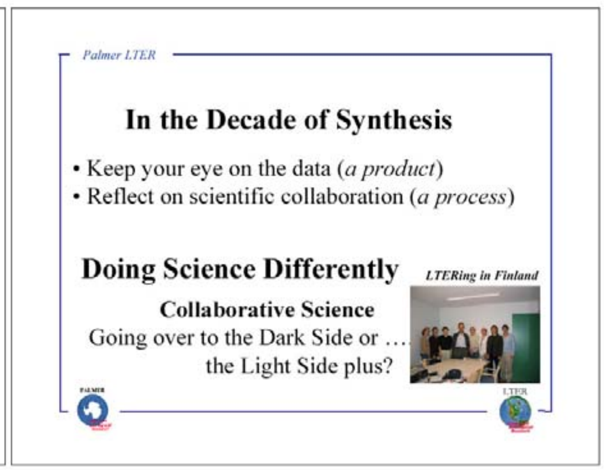
Kiiminki, September 7 2003

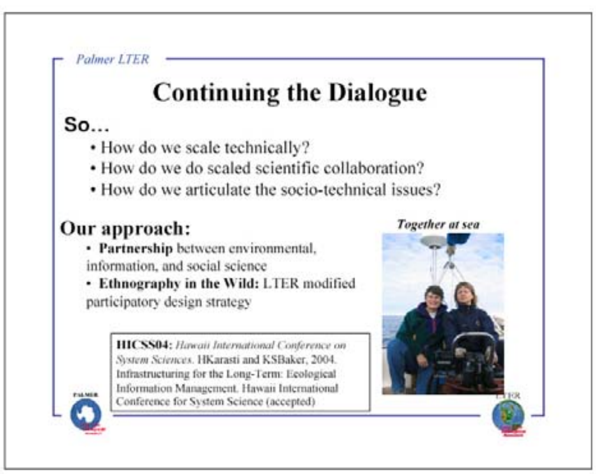
Kelena Karasti, Karen S. Baker and Geoffrey C. Bowker

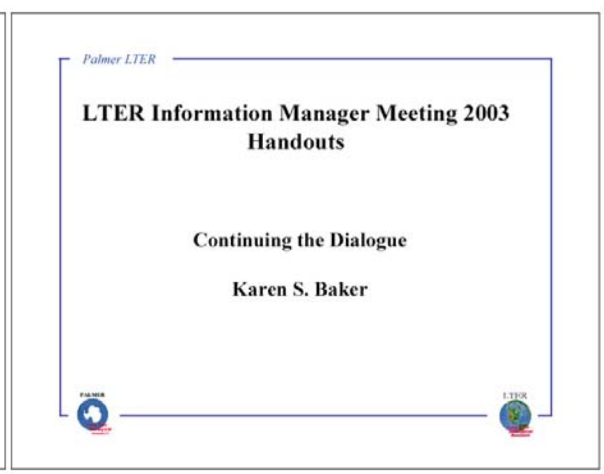
8.14 Appendix: Information Manager Meeting Presentation – Sep 2003



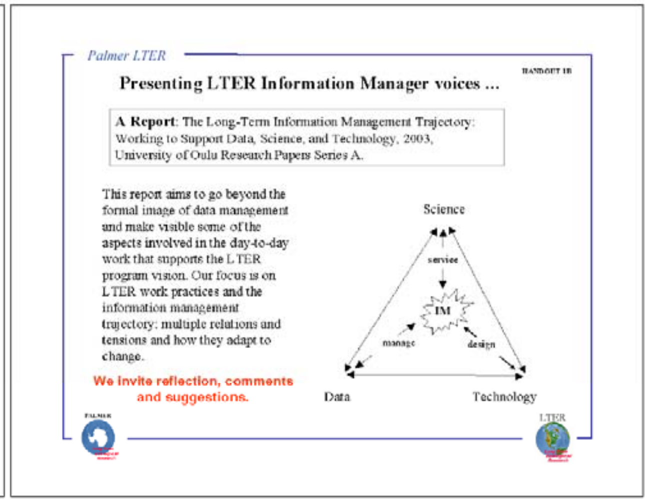


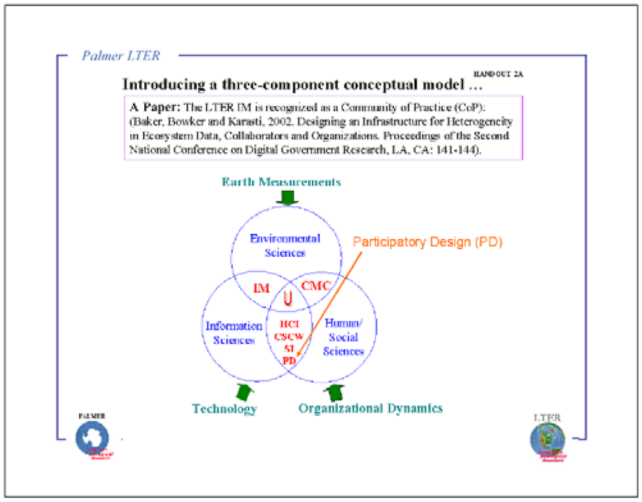


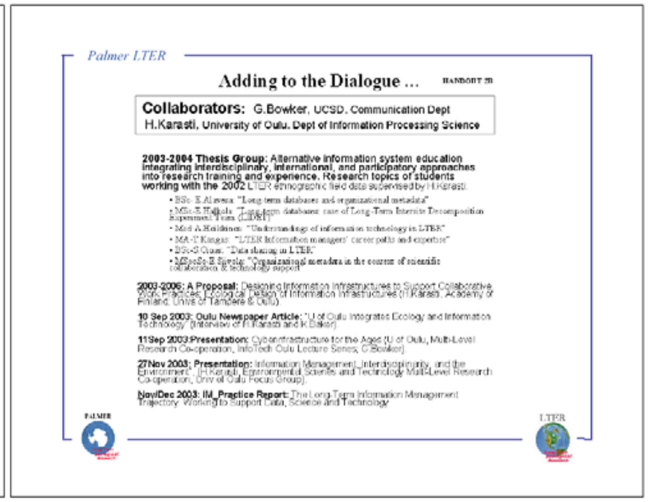





Palmer LTER
Continuing the Dialogue
HANDOUT 1A
Karen Baker, Helena Karasti, Geoffrey Bowker
When both technical and social issues are a recognized part of scientific infrastructuring:
• Informs information management conceptual models and vocabulary
• Contributes to cross domain community dialogues
• Adds multiple perspectives in support of science
• Enables use of design in assessment and information systems
• Creates an ethnographically informed workplace
CSCW02: Computer Supported Cooperative Work. Ecological Storytelling and Collaborative Scientific Activities. H.Karasti, K.Baker, and G.Bowker. ACM SIGGroup Bulletin (in press)
ECSCW03: European Computer Supported Cooperative Work Conference. H.Karasti, K.Baker, and G.C.Bowker (eds), 2003. ECSCW2003 Proceedings of the Computer Supported Scientific Collaboration workshop, Eighth European Conference on Computer Supported Cooperative Work. Helsinki, Finland, 14 September 2003. University of Oulu Department of Information Processing Science Research Papers Series: A34,55pp. ISBN: 951-42-7121-1; ISSN: 0786-8413.
HICSS04: Hawaii International Conference on System Sciences. KSBaker and HKarasti, 2004. Infrastructuring for the Long-Term: Ecological Information Management. Hawaii International Conference for System Science (accepted)
PALMER
LTER


Palmer LTER
HANDOUT 1B
Presenting LTER Information Manager voices ...
A Report: The Long-Term Information Management Trajectory: Working to Support Data, Science, and Technology, 2003, University of Oulu Research Papers Series A.
This report aims to go beyond the formal image of data management and make visible some of the aspects involved in the day-to-day work that supports the LTER program vision. Our focus is on LTER work practices and the information management trajectory: multiple relations and tensions and how they adapt to change.
We invite reflection, comments and suggestions.
Science
service
IM
manage
design
Data
Technology
PALMER
LTER


Palmer LTER
HANDOUT 2A
Introducing a three-component conceptual model ...
A Paper: The LTER IM is recognized as a Community of Practice (CoP): (Baker, Bowker and Karasti, 2002. Designing an Infrastructure for Heterogeneity in Ecosystem Data, Collaborators and Organizations. Proceedings of the Second National Conference on Digital Government Research, LA, CA: 141-144).
Earth Measurements
Environmental Sciences
Participatory Design (PD)
IM
CMC
U
Information Sciences
HCI
CSCW
SI
PD
Human/ Social Sciences
Technology
Organizational Dynamics
PALMER
LTER


Palmer LTER
Adding to the Dialogue ...
HANDOUT 2B
Collaborators: G.Bowker, UCSD. Communication Dept
H.Karasti, University of Oulu. Dept of Information Processing Science
2003-2004 Thesis Group: Alternative information system education integrating interdisciplinary, international, and participatory approaches into research training and experience. Research topics of students working with the 2002 LTER ethnographic field data supervised by H.Karasti:
• MA-T Kangas: "LTER Information managers' career paths and expertise"
• BSc-S Oinas: "Data sharing in LTER"
2003-2006: A Proposal: Designing Information Infrastructures to Support Collaborative Work Practices: Ecological Design of Information Infrastructures (H.Karasti), Academy of Finland; Univs of Tampere & Oulu).
10 Sep 2003: Oulu Newspaper Article: "U of Oulu Integrates Ecology and Information Technology" (Interview of H.Karasti and K.Baker).
11Sep 2003:Presentation: Cyberinfrastructure for the Ages (U of Oulu, Multi-Level Research Co-operation, InfoTech Oulu Lecture Series; G.Bowker).
27Nov 2003: Presentation: Information Management, Interdisciplinarity, and the Environment". (H.Karasti, Environmental Sciences and Technology Multi-Level Research Co-operation, Univ of Oulu Focus Group).
Nov/Dec 2003: IM_Practice Report: The Long-Term Information Management Trajectory: Working to Support Data, Science and Technology
PALMER
LTER

8.15 Appendix: LTER All Scientists Meeting Poster

Long-Term Ecological Research Network
All Scientists Meeting
September 18-21 2003 Seattle, Washington
"Embarking on a Decade of Synthesis"
http://www.lternet.edu/asm/2003

Palmer LTER: Information Flow and Management

Karen Baker, Anna Gold, Frank Sudholt, Helena Karasti, Geoffrey Bowker
Institutional Affiliations: Scripps Institution of Oceanography, University of California San Diego Library, San Diego Supercomputer Center, University of Oulu Department of Information Processing Science, UCSD Communication Department

Abstract: Organizational repositories are being constructed today to address the needs of scientific information management in a digital environment. Given the social aspects of information, building useful information systems requires infrastructures that reflect the unified and expressive relationships of data, documents, people, institutions and partnerships. The Palmer Long-Term Ecological Research (LTER) program information management is working in partnership to explore articulation of the LTER community information management practices and to prototype a co-construction of a low barrier bibliographic referatory/repository.

Just as information flows across electronic boundaries today, this work provides an opportunity to gain experience in crossing the traditional research domains (ecological and marine science, information science and digital library efforts, social science and participatory design approaches). The persistence of infrastructures for long-term scientific projects depends in part on offering low barriers for participation as well as on supporting heterogeneous inputs and outputs.

Project goals include investigation of how short-term/local approaches can be compatible with long-term federation strategies since they are critical to initiating information flow, contributing to knowledge diversity, creating reflexivity in development processes as well as ensuring participant engagement and education. In considering the ramifications of sociotechnical issues and ontological codification on information collections, the importance of the multiple dimensions of design, the varied forms of formal and informal communication, and the nature of tacit and explicit knowledge are highlighted

8.16 Appendix: LTER DataBits Good Reads (1999-2003)

LTER DataBits Good Reads

http://lternet-183.lternet.edu/doc_archive/newsletters/databits/

---

99 Fall
Good Read: Federated Database Vocabulary
- Karen S. Baker, PAL LTER

As we come to terms with networked data systems within the LTER, there is a need for a broadened stock of words to capture and describe the changing conceptual vistas of information management. Sheth and Larson (ACM Computing Surveys 22:183-236) help the community into the 1990's by presenting a pertinent vocabulary with which to address "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases". A range of topics are discussed from semantic heterogeneity to types of federation and architecture schemas for cooperation among independent sites/systems.

99Fall
Noted: Ecologist Outreach at Spokane ESA Meeting
- Karen S. Baker, PAL LTER

The Spokane-Review Newspaper (7August99) carried the story about the ESA annual conference being held for the first time in Spokane this year. It was reported that last year an ESA member was queried "So you're here for that conference, those environmentalists?" and the ESA member had responded with "No! We're scientists!" This interchange alludes to a scientific objectivity expected of ecologists. In fact, the ESA literature defines "ecology is the study of the relationships between living organisms - including human - and their physical environment". It is interesting to note that after this exchange, specifically as an outreach for the Spokane meeting, the society invited nonscientists from the host city to participate as well as discounted education meeting costs for educators. This can be noted as a proactive response to a communications need.

00Spring
Good Reads: How To Manage Data Badly (Parts 1 & 2)
- Darrell Blodgett, BNZ LTER

1.Hale, S.S. 1999. How To Manage Data Badly (Part 1). Bulletin of the Ecological Society of America, 80 (4): pp. 265-268.

(Online PDF file at The Ecological Society of America: pages 20-23).
How To Manage Data Badly (Part 1) is a sarcastic set of ten rules for the database manager or administrator to follow in order to "manage data badly". The rules will probably be recognized by data managers who have seen, or experienced fully or partially compliant systems in the past. Rules such as "Rule 1'.One world, one database" , and "Rule 2. Users are losers" give you a good idea of the humorous content of the article.

2.Hale, S.S. 2000. How To Manage Data Badly (Part 2). Bulletin of the Ecological Society of America, 81 (1): pp. 101-103.

(Online PDF file at The Ecological Society of America: pages 1-3).
How To Manage Data Badly (Part 2) describes rules 11 through 16 which are rules for scientists to insure that their data is managed badly. Again these rules are familiar to data managers who deal the small percentage of scientists who adhere to one or more of these rules. The article concludes by encouraging database managers and scientists to work together following the 16 principles to achieve widespread recognition. The epilogue goes on to describe the importance of good data management and lists several things that are needed to do a better job at managing data.

00Fall
Good Reads: Ecological Data: Design, Management and Processing
- John Briggs, CAP/KNZ LTER

Good Reads: Michener, W.K., and J.W. Brunt eds. 2000. Ecological Data: Design, Management and Processing. Blackwell Science. 180 pp.

This review is very biased. I reviewed this book for the publisher in an earlier version. I liked it then and I still like it. In fact, this book will be a focus of a new course that I am developing next semester (Spring 2001) concerning data management and analysis in ecological science. The editors and the authors have done an excellent job in putting their years of experience into a compact source of knowledge. This is a book every scientist who is associated with a long-term (> 2 years) project should be aware of. More importantly, any person responsible for managing the data of a long-term project should keep this book around them all the time! For most LTER information managers who have been around for a while, attended the annual LTER information manager meetings or who have read most of the material associated with those meetings, you will recognize most of the information presented in this book. If you are in that crowd, don't expect to learn something remarkable and new in this book. However, for the first time you will have a publication that "puts" it all together from thinking about the generation of hypotheses, the nuts and bolts of scientific data bases including archiving to the transformation of numbers into useful information and knowledge. In addition, by reading this book it should help you remind yourself just how important you are to the project. A must-read and have for every LTER information manager.

01Spring
Good Read: Public Access and Use of Electronically Archived Data: Ethical Considerations
- Brent L. Brock, KNZ LTER

Davis, M.A., D. Tilman, S.E. Hobbie, C.L. Lehman, P.B. Reich, J.M. Knops,
S. Naeem, M.E. Ritchie, and D.A. Wedin. 2001. Public Access and Use of
Electronically Archived Data: Ethical Considerations. Bulletin of the Ecological Society of America 82:90-91. This brief article explores the ethical considerations regarding publishing and use of publicly available archived data. The authors stress the need for adoption of a code of ethics to encourage data sharing and protect researchers from inappropriate or unethical use of their data. This article provides food for thought for data managers and ecology researchers alike.

01Spring
Good read: Evolution of a Multisite Network Information System: The LTER Information Management Paradigm
- Wade Sheldon, GCE LTER

Baker, K.S., B.J. Benson, D.L. Henshaw, D. Blodgett, J.H. Porter, and
S.G. Stafford. 2000. Evolution of a Multisite Network Information System:
The LTER Information Management Paradigm. BioScience, 50(11): 963-978.

Once again we have the opportunity to use this space to showcase an outstanding contribution to the field of Ecological informatics by members of our own inner circle. Like 'Ecological Data', reviewed in the Fall 2000 issue of Databits, this paper encapsulates years of LTER information management knowledge and experience in one source. The central theme of the paper is the history and ongoing development of the LTER NIS, presented as a review and case study for application in other research areas. The authors also use the article to educate readers about the basic tenets of scientific information management by providing glossaries of terminology, conceptual diagrams, and lists of core design elements for developing a NIS. This is an inspiring article which should be on every data managers reading list, and should be required reading for any class on scientific information management.

2001 Fall
Good Reads: Ecology Through Time

-Karen Baker, PAL LTER

Kaiser,J, 2001. Ecology Through Time: An Experiment for all Seasons. Science 293(5530):624-627.

In providing an overview of concepts fundamental to the Long-Term Ecological Research network, one can place the recent issue of Science (July 2001) on the reference shelf right next to the Jul/Aug 1990 issue of BioScience which contains that original introductory trio of articles (Swanson and Sparks: Long-Term ecological research and the invisible place; Magnuson: Long-term ecological research and the invisible present; Franklin, Bledsoe, Callahan: Contributions of the long-term ecological research program). This article begins "the NSF's LTER network has proved a smashing success", continues with illustration of the ramifications of network science, and summarizes with site-specific scenarios as well as cross-site tables. Aspects of the spirit of the LTER's practice of science are captured along with examples from the broader community, rounding out this Science issue dedicated to ecosystem science.

2001 Fall
Good Reads: Biodiversity Datadiversity
-Karen Baker, PAL LTER

Bowker, GC, 2000. Biodiversity Datadiversity. Social Studies of Science 30(5): 643-683.

One might consider this article by Bowker an ecological continuation of a conversational thread developed by Brooks in his book the Mythical Man Month (1975). Acknowledging the difficulties arising from the multiplicative factor inherent in team project communications, Bowker strides on into the territory of well-counted megafauna and under-valued microscopic entities to discuss the diversity in databases and transitions in frameworks. This work is a call for social scientists to consider potential contributions given their experience with naming and classifying, context and integration, organizational practice and scientific history. Attention to infrastructure layering impacts first the movement from data discovery to data management and ultimately the communication of knowledge through semantic synthesis. The challenges are confronted as Bowker articulates the nontrivial nature of the informatics task and summarizes: "The information collection effort that is being mounted worldwide is indeed heroic."

2002 Spring
Good Read: Managing Scientific Metadata
- Karen Baker, PAL LTER

Matthew B Jones, Chad Berkley, Jivka Bojilova, and Mark Schildhauer, 2001. Managing Scientific Metadata. IEEE Internet Computing Sep-Oct 2001. (http://computer.org/internet)

This article provides both background into the use of metadata as a method to manage heterogeneous scientific data, summarizing original motivatations, general approaches and community specifics. To cover such ground, the authors adopt an approach that may be regarded as a metaphor reflective of the cross-domain integrative strategies required for information management today. The metaphor, a conversation-within-a-conversation sidebar, highlights metadata standards, XML databases and interoperability issues. An LTER IM Metadata Workshop organized by P.McCarthy (CAP) and supported by the LTER Network Office provided a mechanism for LTER sites to focus on these issues and to consider site metadata transformation strategies compatible with interoperable formats, eg the Ecological Metadata Language. Given LTER Information Management discussions and metadata momentum, this paper provides insight into the broader context.

Good Read: Is It Time to Bury the Ecosystem Concept?
- Maria Vernet and Karen Baker, PAL LTER

Robert O'Neill, Is it time to bury the ecosystem concept? Ecology 82(12), 2001 pp3275-3284

O'Neill reminds the reader that the ecosystem approach is a model we impose on nature and examines model assumptions including stability, fixed boundaries and homogeneity. Use of the ecosystem model over the past decades has revealed some aspects of natural systems are better explored through alternative methods. One focus is on the importance and impact of local stability and on the equally important concept of long-term sustainability. His framework provides the bigger context, beyond basin to global, so can address questions such as 'what is left in a location after ecosystem flight occurs'.

The statement "Homo sapiens is not an external disturbance, it is a keystone species within the system" suggests that with ecosystem defining characteristics like dispersal range (which in case of dogs and humans is global), we discover the disturbance/recovery of 'spaceship earth' is indeed in the hands of the species with a capacity for

a) information integration
b) organizational memory
c) humor
d) all of the above
e) none of the above

The answer will be forthcoming.

2002 Fall
Good Read: Ecological Vignettes
-Karen Baker, PAL LTER

Eugene Odum, 1998. Ecological Vignettes: Ecological Approaches to Dealing with Human Predicaments, Harwood Academic Publishers, 269p.

Eugene Odum's 'Ecological Vignettes' brings to mind Rachel Carson's 'Silent Spring' in that it presents the particular along with some general ramifications. The book is divided into two parts with vignettes followed by more detailed essays that reference the scientific literature. The vignettes synthesize insights from a broad ecological career that has focused over the years on local as well as global systems with some sensitivity to political, economic, and social ramifications. The scaling of knowledge grounded by experience with watershed studies provides a much needed articulation about issues at the ecosphere level. Odum builds from a starting point, his determining factor, that the human population has reached the maximum carrying capacity of the earth as a whole. With a nontraditional, multi-tier presentation format, Odum invites and then supports participation from a broad audience by providing tools in the form of access to information in an approachable format. From the dark side of technology to the tyranny of small decisions, the book provides an often elusive bigger picture relevant to individual reflection as well as national action.

Good Read: A History of the Ecosystem Concept in Ecology
- Karen Baker, PAL LTER

Frank Golley, 1993. A History of the Ecosystem Concept in Ecology, Yale University Press, 254p.

In order to gain insight into a concept, context is provided by some often-nonlinear historical events. As LTER community members, we benefit from the the sweep and the depth of Frank Golley's presentation on ecosystem science. As information managers, we
benefit from his recognition of the need for information management in combining, extending and passing on the data that science gathers. One historical note worth mentioning because it highlights an important distinction sometimes lost in the tacit understanding of our current research environment, is that LTER is not an acronym for Long-Term Ecosystem Research (p118) but rather for Long-Term Ecological Research. Is this an important distinction? Golley provides organizational examples, contrasting the business model for big science programs with a more academic approach. LTER, as a network, is a community organization model that explicitly adopts an integrative embrace of ecology, avoiding potential misunderstandings over the multiple levels of meaning and history associated with the term 'ecosystem research'. The LTER understaking is an ongoing rebalance of understandings generated by the multiple views afforded by the spectrum of reductionist to holistic, by the elements juxtapositioned with the whole.

2003 Spring

Good Read: Information Ecology
-Karen Baker, PAL LTER

Davenport, T.H. Information Ecology: Mastering the Information and Knowledge Environment. Oxford University Press, 1997.

Davenport in Chapter 1 makes clear his views on information management through presentation of a pair of lists that invite the reader to compare and contrast. He lists four beliefs of those looking to technology to solve our information challenges:

*information is easily stored n computers - as 'data';
*modeling computer databases is the only way to master information complexity;
*information must be common throughout an organization;
*technology change will improve the information environment

and four beliefs of those taking a more ecological approach to information management:

*information is not easily stored on computers-and is not 'data'
*the more complex an information model, the less useful it will be;
*information can take on many meanings in an organization;
*technology is only one component of the information environment

The brief lists are an effective method to highlight differences between technological and sociotechnical approaches. Be wary browsing this author on the bookstore shelf as there is a Thomas O. Davenport who has written a book "Human Capital: What It Is and Why people Invest in It" (1999) that details a popular contemporary mangement philosophy. It's interesting but distinct from Thomas H.Davenport's "Information Ecology".

Good Read: The Invisible Present
-Karen Baker, PAL LTER

Magnuson, John, "The Invisible Present" in Ecological Time Series. T.M. Powell and J.H.Steele (eds), 1995.

A classic LTER article providing insight into the multitude of views derived from varying temporal time scales in ecological science. A series of graphs, showing the changing view resulting from opening up of a time series, illustrates a point fundamental
to the philosophy of LTER. Although published earlier in BioScience as one of a trio of LTER articles (1990), this adaption of the article becomes part of a broader context when it appears within an ecological time series book that spans the land, ocean and human health domains.

2003 Fall

Good Read: BioScience January 2003 Special Issue
- Karen Baker, Palmer Station

John E. Hobbie (ed), 2003. A Special Section on the US Long Term Ecological
Research Network. BioScience 53(1).

A special section in Bioscience about LTER provides a comprehensive historical context for long-term research from pre International Biological Program (IBP) time through the work of LTER today. Two articles summarizing the LTER program and its accomplishments are followed by six articles on cross-site research topics: climate forcing, land-use, biodiversity, system disturbance, system variability, and mechanistic modeling. The collection of articles presents the LTER community, highlighting its mission and selected ecological the rich intellectual and data resources, the series gives insight both to the research community approach as well as to the long-term data legacy. The overview offers an opportunity to consider the LTER process as a whole.

Good Read: Steps Towards an Ecology of Infrastructure
- Karen Baker, Palmer Station & Helena Karasti, Univ of Oulu Dept of Information Processing Science

Susan Leigh Star and Karen Ruhleder, 1994. Steps Towards an Ecology of Infrastructure: Complex problems in design and access for large-scale collective systems. In Transcending Boundaries: Proceedings of the conference on Computer Supported Cooperative Work (CSCWb 94), 22-26 October, Chapel Hill, NC. ACM Press, New York, p. 253-264.

The LTER represents one model of a networked community organization. The Worm Community System (WCS), a distributed software environment with successes and challenges similar to and different from LTER, represents another collaboratory model. Star and Ruhleder (1994), using ethnographic methods to conduct research on the WCS, present an analytic framework using multiple levels of understandings to capture this community's not-so-well-structured structures and not-so-well-expressed tensions.

The traditional "wires and pipes" infrastructure metaphor is broadened to encompass relationships with system users and change within organizations. Such multidimensional views of infrastructure are important to collaboratory participants facing choices about how technology standards will be used to support site science, how technology decisions will influence data practices, and how design approaches influence information system development. This work describes explicitly how local practices interface with large-scale structure, demonstrating how local customization is in tension with the development of common standards. The complexity and interdependence of everyday work practices are found to be critical elements when considering both technical and sociotechnical challenges.

8.17 Appendix: Information Management and System Venues

Information Management and Information System Conferences, Journals, Proceedings, and Bulletins

Journal Rankings:

Whitman and Hendrickson:
http://www-users.cs.york.ac.uk/~kimble/research/Ranks.html

Alba Index of IS journals (The Athens Laboratory of Business Administration): Communications of the ACM 44(9): 29-33
http://www.alba.edu.gr/survey/IS/ISJournals.pdf

Walstrom and Hardgrave, 2001: Forums for information systems scholars:III. Information and Management 39: 117-124.
http://lamp.infosys.deakin.edu.au/journals/IM.39.2.117.124.pdf
http://lamp.infosys.deakin.edu.au/journals/index.php

Information Systems Publications: A Classification, 1995
http://www.anu.edu.au/people/Roger.Clarke/ISRes/INFSRefs.html

Summaries of Summaries:

Index of Information Systems Journals (2003)
http://lamp.infosys.deakin.edu.au/journals/index.php

Electronic Journals Information Systems
http://www.cityu.edu.hk/lib/eres/ej/subject/ej_infosys.htm

Information Science related Academic Journals
http://uib.no/uraadet/nfiv/reltid.htm#Journals

Cross-comparison of six rankings (Mylonopoulos & Theoharakis, 2001; Whitman et al., 1999; Hardgrave & Walstrom, 1997; Walstrom et al., 1995; Holsapple et al., 1994; Gillenson & Stutz, 1991):
http://www.bus.ucf.edu/csaunders/newjournal.htm

Calendar Summary

International Calendar of Information Science Conferences (ICISC)
http://icisc.neasist.org

Conferences, Proceedings, Journals, and Bulletins

ACM
NAME: ***Association of Computing Machinery***
URL: http://www.acm.org

The *Association of Computing Machinery* is an organization funded in 1947 to advance the skills of information technology professionals and students worldwide. ACM includes several special interest groups (SIGs) such as Groupware (SIGGROUP) and Management of Data (SIGMOD) that hold periodic meetings and publish journals, e.g. SIGGROUP Bulletin.

SIGGROUP
The ACM Special Interest Group on Supporting Group Work (SIGGROUP, formerly called SIGOIS) is interested in topics related to computer-based systems that have a team or group impact in workplace settings. A strong emphasis of SIGGROUP is the integration of multiple computer-based tools and technologies and the impact on the human activities supported by those tools and technologies. Relevant issues include design, implementation, deployment, evaluation, methodologies and impact that arise when researching computer-based systems in a development environment. SIGGROUP publishes the SIGGROUP Bulletin three times a year.

SIGMOD
The ACM Special Interest Group on Management of Data investigates the development and application of database technology on a full range of computer organizations. The scope of interests and members is wide with an almost equal mix of people from industry and from academia. SIGMOD sponsors an annual conference that is regarded as one of the most important in the field, particularly for practitioners. Similarly, the quarterly newsletter SIGMOD Record is a valuable resource for state-of-the-art information and includes the annual conference proceedings.

ACM SIGCHI
NAME: ***Association of Computing Machinery Special Interest Groups Computer Human Interaction***
URL:http://www.acm.org/sigchi/
The *Association of Computing Machinery* (ACM) *Special Interest Groups Computer Human Interaction* (SIGCHI*)* brings together people working on the design, evaluation, implementation, and study of interactive computing systems for human use. ACM SIGCHI provides an international, interdisciplinary forum for the exchange of ideas about the field of human-computer interaction (HCI).

ASIS&T
NAME: ***American Society for Information Science and Technology***
URL: http://www.asis.org/
URL: http://www.asis.org/AboutASIS/asis-sigs.html
URL: http://www.asis.org/Conferences/AM04
The American Society for Information Science and Technology brings together those interested in a wide variety of information and technology topics. ASIST is a professional society that bridges the gap between diverse needs of researchers, developers and end users and between the challenge associated with emerging technologies and applications ranging across the fields of library and information science, communication, networking technologies, and computer science. ASIS&T holds an annual conference with a proceedings. A themed Bulletin is published bimonthly. There are a series of special interest groups (SIGs) including ones on Human-Computer Interaction (HCI), Information Architecture (IA), Digital Libraries (DL), and Scientific & Technical Information Systems (STI) or Social Informatics (SI).

ASIS&T SIG AI
NAME: ***American Society for Information Science and Technology Information Architecture***
URL: http://www.iasummit.org
Information Architecture Summit. The Information Architecture summit is an emerging discipline and a key player in web, multimedia, software, and even product design. Holds an annual meeting since 2000 with a proceedings.

CSCW
NAME: ***Computer Supported Cooperative Work***
URL: http://www.acm.org/conferences/cscw2002/
The *Computer Supported Cooperative Work* (CSCW) is a computer systems' research and development community that brings together the social and technical aspects for supporting collaboration. CSCW is a multidisciplinary area of research devoted to the use of computers to support cooperative work. The community encompasses interests in systems development and design, the theory and practice of CSCW, the installation and use of CSCW systems, the application of novel technologies, and the relation between the social and the technical. There is a meeting Proceedings.

ECSCW
NAME: ***European Computer Supported Cooperative Work***
URL: http://www.cti.dtu.dk/projects/cscw/ECSCW.html
URL: http://www.itu.dk/people/schmidt/ecscw.html
The European Conference on Computer Supported Cooperative Work is sponsored by the European CSCW Foundation, an association of European CSCW research sites. Proceedings exist since 1991.

CSCW-JCC
NAME: ***The Journal of Collaborative Computing***
URL: http://www.kluweronline.com/issn/0925-9724/contents
Computer Supported Cooperative Work (CSCW): The Journal of Collaborative Computing is devoted to innovative research in computer supported cooperative work (CSCW). It provides an interdisciplinary forum for the debate and exchange of ideas concerning theoretical, practical, technical, and social issues in CSCW. The journal arose as a response to the growing interest in the design, implementation and use of technical systems (including computing, information, and communications technologies) which support people working cooperatively. The scope of the CSCW journal remains to encompass the multifarious aspects of research within CSCW and related areas – from ethnographic studies of cooperative work to reports of the development of CSCW systems and their technological foundations.

PDC
NAME: Participatory Design Conference
URL: http://cpsr.org/conferences/
Participatory Design Conference is an international forum where this emerging community can meet, exchange ideas and experiences, and investigate the incorporation of participatory design approaches in new areas such as: product development, long-term system maintenance and redesign, and settings in the developing world. PDC is sponsored by Computer Professionals for Social Responsibility (CPSR).

GRID Forum
NAME: Grid Computing
URL: http://grid.org
URL: http://www.globalgridforum.com
URL: http://www.sdsc.edu/GridForum (annual meeting)
Grid computing is a form of distributed computing that involves coordinating and sharing computing, application, data, storage, or network resources across dynamic and geographically dispersed organizations. Grid technologies promise to change the way organizations tackle complex computational problems. However, the vision of large scale resource sharing is not yet a reality in many areas. Grid computing is an evolving area of computing, where standards and technology are still being developed to enable this new paradigm.

HICSS
NAME: *Hawaii International Conference for System Sciences*

URL: http://www.hicss.org
The *Hawaii International Conference for System Sciences*,The IEEE Computer Society and the ACM are among the sponsors of HICSS annual conference. Many conferences focus on a specific discipline or subject. Although specialization is important, HICSS has chosen to become one of the few general-purpose conferences addressing issues in the areas of computer science, computer engineering, and information systems. The fundamental purpose of this conference is to provide a forum for the exchange of ideas, research results, development activities, and applications. HICSS brings together highly-qualified interdisciplinary professionals in an interactive environment. An annual proceedings is published.

ISSDBM
NAME: ***International Conference on Scientific and Statistical Database Management***
URL: http://www.informatik.uni-trier.de/~ley/db/conf/ssdbm/
This international conference will bring together scientific domain experts, database researchers, practitioners and developers for the presentation and exchange of current research on concepts, tools and techniques for scientific and statistical database applications. There is an annual conference and proceedings.

IEEE Meta-Data Conference
NAME: ***Institute of Electrical and Electronic Engineers*** **(IEEE)** ***Meta-Data Conference***
URL: http://www.llnl.gov/liv_comp/metadata
The objectives of the *Institute of Electrical and Electronic Engineers* (IEEE) *Meta-Data Conference* series of workshops and conferences are to provide a forum to address meta-data issues, bring the different communities together for technical interchange, hear the various perspectives and facilitate development and usage of meta-data.

JCDL
NAME: Joint Conference on Digital Libraries
URL: http://www.jcdl.org
URL: http://www.jcdl.org/past-event-conf.shtml
Joint Conference on Digital Libraries has a proceedings. It is a major international forum focusing on digital libraries and associated technical, practical, and social issues. JCDL encompasses the many meanings of the term "digital libraries", including (but not limited to) new forms of information institutions; operational information systems with all types of digital content; new means of selecting, collecting, organizing, and distributing digital content; digital preservation and archiving; and theoretical models of information media, including document genres and electronic publishing. JCDL holds an annual meeting and publishes a proceedings. There are a series of Digital Library. JCDL combines the DL:ACM Conference on Digital Libraries and the ECDL: European Conference on Research and Advanced Technology for Digital Libraries

JASIST
NAME: ***Journal of the American Society for Information Science and Technology***
URL: http://www.asis.org/Publications/JASIS/jasis.html
This journal serves as a forum for new research in information transfer and communication processes in general, and in the context of recorded knowledge in particular. Concerns include the generation, recording, distribution, storage, representation, retrieval, and dissemination of information, as well as its social impact and management of information agencies. There is a strong emphasis on new information technologies and methodologies in text analysis, computer based retrieval systems, measures of effectiveness, and the search for patterns and regularities in measures of existing communication systems. The orientation is toward quantitative experimental work, but significant qualitative and historical research is also welcome. Perspectives, a journal within the journal, containing collections of papers on single topics normally operationally practical in nature, is often included. Special topic issues are also often seen.

JEI
NAME: ***Journal of Environmental Informatics***
URL: http://www.iseis.org/jei-subs.htm
Journal of Environmental Informatics (JEI) is an international, peer-reviewed, and interdisciplinary publication designed to foster innovative research on systems science and information technology for environmental management. The journal aims to motivate and enhance the integration of environmental information and systems analysis to help develop management solutions that are consensus-oriented, risk-informed, scientifically-based and

cost-effective. The topics addressed by the journal include: decision and risk analysis for environmental management, mathematical methods for systems modeling and optimization, environmental data and information management, ecological modeling and simulation, GIS, RS, geographical analysis, expert systems, environmental systems science and information technology

JoDI
NAME: ***Journal of Digital Information Management***
URL: http://jodi.tamu.edu
Journal o f Digital Information Management (JoDI) is a journal sponsored by the Digital Information Research Foundation. It is a quarterly journal in digital information science and technology, concentrating on all aspects of digital information management. It broadly covers digital information processing, digital content management, digital world structuring, digital libraries, metadata, information management and other related fields. It is an international peer reviewed journals containing original research papers, ongoing research, technology, reviews, reports of progress, short notes, and forthcoming events.

NBII - TBP eForum
NAME: National Biological Information Infrastructure Towards Best Practices
URL: http://www.nbii.gov
URL: http://www.nbii.gov/datainfo/bestpractices/index.html
National Biological Information Infrastructure. The National Biological Information Infrastructure (NBII) is a broad, collaborative program to provide increased access to data and information on the nation's biological resources. The NBII links diverse, high-quality biological databases, information products, and analytical tools maintained by NBII partners and other contributors in government agencies, academic institutions, non-government organizations, and private industry. NBII partners and collaborators also work on new standards, tools, and technologies that make it easier to find, integrate, and apply biological resources information. Resource managers, scientists, educators, and the general public use the NBII to answer a wide range of questions related to the management, use, or conservation of this nation's biological resources.

The Towards Best Practices (TBP) eForum is a Web-based resource designed for those involved in studying and managing the complex interactions between life forms - including human populations - and the environment. This free archive and public forum allowing users to engage in moderated debates of submitted Best Practices

VLDB
Name: ***Very Large Database Conference***
URL: http://www.vldb.org/
Very Large Data Base Endowment Inc. (VLDB Endowment) is non-profit organisation incorporated in the United States for the sole purpose of promoting and exchanging scholarly work in databases and related fields throughout the world. There is an annual conference, an annual proceedings, and a quarterly VLDB Journal.

Other Information System Journals and Conferences

4S
NAME: ***Society for Social Studies of Science***
URL: http://www.4sconference.org/
URL: http://www.jstor.org/journals/07380526.html
The Society for Social Studies of Science (4S) is a nonprofit, professional association founded in 1975 and now has an international membership. The main purpose of 4S is to bring together those interested in understanding science, technology, and medicine, including the way they develop and interact with their social contexts Conference papers on all subjects connected to the social and cultural analysis of science, technology, and medicine.

AJIS
NAME: ***Australian Journal of Information Systems***
URL: http://www.uow.edu.au/ajis/ajis.html
The Australian Journal of Information Systems (AJIS) is a refereed journal that publishes articles (twice per year) contributing to Information Systems theory and practice.

BNCOD
NAME: ***British National Conference on Databases***
URL: http://www.shef.ac.uk/bncod2002/
URL: http://www.macs.hw.ac.uk/BNCOD21/index.htm
URL: http://www.macs.hw.ac.uk/BNCOD21/
British National Conference on Databases (BNCOD)(Edinburough/subtrack n CSCW)
For the past twenty years BNCOD has attracted an International audience to discuss the leading research topics of the day in the field of Data, Knowledge and Information systems. Leading research from the UK and from throughout the world has been presented at the conference, and many fruitful and lasting collaborations have been created.

C&T
NAME: *International Conference on Communities and Technologies*
URL: http://www.iisi.de/114.0.html?&L=3
International Conference on Communities and Technologies considers the relationship between communities and technology as a topic of major research interest. 'C&T' conference serves as a forum for stimulating and disseminating research into all facets of communities and information technology. The nature of the field requires multidisciplinary research efforts involving researchers from different fields of applied computer science (Computer Supported Cooperative Work, Computer Supported Collaborative Learning, Artificial Intelligence, Information Retrieval, Human Computer Interaction, Information Systems) and social sciences (Economics, Management Science, Psychology, Political Science, Sociology, Ethnography, Discourse Analysis).

COTS
NAME: ***International Conference on Commercial Off-the-Shelf-Based Software Systems***
URL: http://www.iccbss.org
International Conference on Commercial Off-the-Shelf-Based Software Systems (ICCBSS 2004) focuses on exchanging ideas about current best practices and promising research direction in creating and maintaning systems that incorporate COTS software products. Successful integration of commercial off-the-shelf (COTS) software in support of key business, operational, or mission related processes has become essential to creating and maintaining responsive systems that keep pace with technology advances. ICCBSS focuses on exchanging ideas about current best practices and promising research directions in creating and maintaining systems that incoporate COTS software products. It provides a forum in which researchers and practitioners from industry, government, and universities can gather to exchange ideas and results. Proceedings exist from 2002 and 2003.

EJIS
NAME: ***European Journal of Information Systems***
URL: www.palgrave-journals.com/ejis
European Journal of Information Systems is a new journal with a scope of providing a perspective n the theory and practice of information systems for a global audience.

IADIS
NAME: ***International Association for Development of the Information* Society**
URL: http://www.iadis.org/wbc2004
Web based communities 2004 and Applied Computing 2004 International Conferences. The Web Based Communities 2004 Conference aims at bringing together new vital understanding of WWW communities and what new initiatives mean. Each new perspective is potentially a catalyst for finding new architectures. National and regional-oriented communities may soon be relegated to a subordinate position compared to interest-oriented communities. Multiculturalism, critical thinking, expressing aesthetic aspects of our identity, and finding sparring partners for sharpening our ideologies, are all processes that need the new communication infrastructures.

ICIS
NAME: International Conference on Information Systems
URL: http://business.queensu.ca/icis
The proceedings of the International Conference on Information Systems are published annually. This is a conference of the The Association for Information Systems (AIS)

IR
NAME: ***Information Research***
URL: http://informationr.net/ir
Information Research (IR) is a *freely available*, international, scholarly journal, dedicated to making accessible the results of research across a wide range of information-related disciplines. It is privately published.

ISJ
NAME: ***Information Systems Journal***
URL: http://www.blackwellpublishing.com/journal.asp?ref=1350-1917
The *Information Systems Journal (ISJ)* aims to promote the study of, and interest in, information systems and to publish articles that reflect the wide and interdisciplinary nature of the subject. Articles are welcome on research, practice, experience, current issues and debates. The journal seeks to integrate technological disciplines with management and other

ISR/INFORMS
NAME: Information Systems Research
URL: http://isr.katz.pitt.edu/
URL: http://www.informs.org/Pubs
Information Systems Research (ISR) is a journal of INFORMS, the Institute for Operations Research and the Management Sciences. ISR is a leading international journal of theory, research, and intellectual development, focused on information systems in organizations, institutions, the economy, and society.

IT&Society
NAME: Information Technology and Society
URL: http://www.stanford.edu/group/siqss/itandsociety/v01i03/
A web journal published quarterly studying how technology affects society.

JAIS
NAME: Journal of the Association for Information Systems
URL: http://jais.isworld.org/
*Journal of the Association for Information Systems (*JAIS) is inclusive in scope and covers all aspects of Information Systems and Information Technology. The Journal publishes rigorously developed and forward looking conceptual and empirical contributions. The Journal encourages multidisciplinary and nontraditional approaches.

JDIM
NAME: Journal of Digital Information Management
URL: http://www.dirf.org/jdim/
Journal of Digital Information Management works to identify the optimum strategy and best practice for the ideal digital information management are the goals of this scholarly journal. JDIM would maintain its commitment to excellence

JISE
NAME: ***Journal of Information Science and Engineering***
URL: http://www.iis.sinica.edu.tw/page/jise/Introduction.html
Journal of Information Science and Engineering. The purpose of this Journal is to provide an open forum to publish high quality research papers in the areas of information science and engineering to promote the exchange of research ideas, experiences and results. It is dedicated to the dissemination of information

JIT
NAME: ***Journal of Information Technology***
URL: http://www.palgrave-journals.com/jit
Journal of Information Technology: Organization, Management, Information and Systems re-launches itself this year with a focus on 'technochange', that change where technology plays a large part.

JSTOR
NAME: ***Journal of Information for Social Studies of Science***
URL: http://www.jstor.org/journals/03063127.html
In the broadest sense, JSTOR's mission is to help the scholarly community take advantage of advances in information technologies.

MISQ
NAME: ***Management of Information Systems Quarterly***
URL: http://www.misq.org
Management of Information Systems Quarterly (MISQ) publishes research concerning both the management of information technology and the use of information technology for managerial and organizational purposes.
and scholarship through the publication of scholarly contributions

SCI
NAME: ***World Multi-Conference on Systemics, Cybernetics and Informatics***
URL: http://www.iiis.org/sci2004/website/default.asp

SCI conferences are international forums for scientists and engineers, researchers and, consultants, theoreticians and practitioners in the fields of Systemics, Cybernetics and Informatics. It is a forum for focusing into specific disciplinary research, as well as for multi, inter and trans-disciplinary studies and projects. One of its aims is to relate disciplines fostering analogical thinking and, hence, producing input to the logical thinking.

SJIS
NAME: ***Scandinavian Journal of Information Systems***
URL: http://www.cs.auc.dk/~sjis
The *Scandinavian Journal of Information Systems* (SJIS) is the journal of the IRIS (Information systems research seminar in Scandinavia) association. The roots of the journal can be found in the tradition of annual IRIS conference as an annual working seminar for Scandinavian researchers and PhD students. Over the years we have been happy to see an increasing number of participants from outside Scandinavia. After its first few meetings in Finland, the seminar now alternates between the Scandinavian countries.

TIS
NAME: ***The Information Society***
URL: http://www.indiana.edu/~tisj/
The Information Society (TIS) journal, published since 1981, is a key critical forum for leading edge analysis of the impacts, policies, system concepts, and methodologies related to information technologies and changes in society and culture. Some of the key information technologies include computers and telecommunications; the sites of social change include homelife, workplaces, schools, communities and diverse organizations, as well as new social forms in cyberspace.
on computer science, computer engineering, and computer systems.
areas such as psychology, philosophy, semiology and sociology.

TIMJ
NAME: ***The Information Management Journal***
URL: http://www.arma.org/publications/journal/index.cfm
The Information Management Journal is the professional journal of *The Association for International Management Professionals* (ARMA International). Its purpose is to advance the field's knowledge base and help satisfy the growing information, educational, and intellectual needs and interests of information management professionals.

8.18 Appendix: Project Publications

Baker, K.S., A.K.Gold, and F.Sudholt, 2003. FLOW: Co-constructing Low Barrier Repository Infrastructure in Support of Heterogeneous Knowledge Collections(s). Proceedings of the third ACM/IEEE-CS JCDL.

Baker, K., J.Brunt and D.Blankman, 2002. Organizational Informatics: A Research Network Site Description Directory Design. Proceedings of the SCI2002 6th World Multi-Conference on Systematics, Cybernetics and Informatics.

Baker, K.S., 2004. 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, pp. 5-7.

Bowker, G. and K.Baker, (accepted). Information Ecology: Open System Environment for Data, Memories and Knowing. Journal of Intelligent Information Systems. BDEI Special Series.

Brunt, J., P.McCartney, K.Baker and S.Stafford, 2002. The Future of Ecoinformatics in Long Term Ecological Research. Proceedings of the SCI2002 6th World Multi-Conference on Systematics, Cybernetics and Informatics

Halkola, E., 2004. Scientific Collaboration and Information Infrastructures – Information Practices in LIDET-experiment. IRIS 27 August 14-17, 2004, Falkenberg, Sweden, http://www.hh.se/iris2; http://w3.msi.vxu.se/users/per/IRIS27/iris27-1107.pdf

Jackson, S.J. and K. S. Baker, 2004. Ecological Design, Collaborative Care, and Ocean Informatics, Proceedings of the Participatory Design Conference, Toronto.

Karasti, H., KSBaker and GCBowker. 2002. Ecological Storytelling and Collaborative Scientific Activities, (SIGGROUP Bulletin 23(2):29-30.

Karasti, H., KSBaker and GCBowker, 2004. ECSCW 2003 Computer Supported Scientific Collaboration (CSSC) workshop report. SIGGROUP Bulletin 24(1).

Karasti, H. K. Baker, and G.C. Bowker 2003. ECSCW2003 Proceedings of the Computer Supported Scientific Collaboration Workshop, Eighth European Conference on Computer Supported Cooperative Work. Helsinki, Finland, 14 September 2003, University of Oulu Department of Information Processing Science Research Paper Series. A34, 55pp. ISBN: 951-2-7121-1; ISSN:0786-8413.

Karasti, H. and K. Baker, 2004. Infrastructuring for the Long-Term: Ecological Information Management. Proceedings of the Hawaii International Conference on System Sciences (HICSS) 5-8 January 2004, Big Island Hawaii. New Brunswick, NJ, IEEE.

Karasti, H. and A-L. Syrjanen. 2004. Artful Infrastructuring in Two Cases of Community PD. Proceedings of the Participatory Design Conference, Toronto.

Melendez, E. and K. Baker, 2002. Common Information Management Framework: in Practice. Proceedings of the SCI2002 6th World Multi-Conference on Systematics, Cybernetics and Informatics.