Computer Supported Cooperative Work (2006) 15:321–358
DOI 10.1007/s10606-006-9023-2

Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network

HELENA KARASTI[1,]*, KAREN S. BAKER[2] & EIJA HALKOLA[1]
[1]*Department of Information Processing Science, University of Oulu, PO Box 3000, FIN-90014, Oulu, Finland (Phone: +358-8-553-1913; Fax: +358-8-553-1890; E-mail: helena.karasti@oulu.fi);* [2]*Scripps Institution of Oceanography, University of California, San Diego, 8615 Discovery Way; 2252 Sverdrup Hall, La Jolla, CA, 92093-0218, USA*

Abstract. This paper aims to enrich the current understanding of data curation prevalent in e-Science by drawing on an ethnographic study of one of the longest-running efforts at long-term consistent data collection with open data sharing in an environment of interdisciplinary collaboration. In such a context we identify a set of salient characteristics of ecological research and data that shape the data stewardship approach of the Long Term Ecological Research (LTER) network. We describe the actual practices through which LTER information managers attend to the extended temporal scale of long-term research and data sets both through data care work and information infrastructure development. We discuss the issues of long-term and continuity that represent central challenges for data curation and stewardship. We argue for more efforts to be directed to understanding what is at stake with a long-term perspective and differing temporal scales as well as to studying actual practices of data curation and stewardship in order to provide more coherent understandings of e-Science solutions and technologies.

Key words: cyberinfrastructure, data stewardship, information management, ecology, long-term perspective, scientific collaboration

1. Introduction

Data issues of scientific collaboration are a key topic at the intersection of Computer Supported Cooperative Work (CSCW) and e-Science. The vision of e-Science (or Cyberinfrastructure) with interest in large-scale science carried out through distributed global collaborations brings forward the issues of access to and sharing of scientific data collections together with the supporting technologies of networks and computing resources (Atkins et al., 2003; UK Research Council e-Science definition, 2001). In CSCW, in turn, collaborative data and information practices have been at the heart of research and technology development (e.g. Greif and Sarin, 1987), albeit with science practices as only one domain of interest and more typically associated with smaller scale collaborative settings (e.g. Birnholtz and Bietz, 2003;

Sonnenwald, 2003; Chin and Lansing, 2004). The particular data issue of interest in this paper, namely data curation, has only recently been put on the e-Science agenda (Hedstrom, 2003; Lord and Macdonald, 2003; Hodge and Frangakis, 2004; National Science Board, 2005). Data curation, embracing "the care of the record within scientific context and environment" (Lord and Macdonald, 2003, p. 45), is complementary as well as critical in providing a substrate for the successful access, sharing and (re)use of data collections, issues that already have received extensive attention (e.g. Sterling and Weinkam, 1990, Hilgartner, 1995; Van House et al., 1998; Helly et al., 2002; Newman et al., 2003; Arzberger et al., 2003; Zimmerman, 2003; Jirotka et al., 2005).

The recognized importance of data curation in e-Science relates to an awareness of the exponentially increasing volumes of primary data in digital form generated by automated collection of data through "next generation" experiments, simulations, sensors and satellites (e.g. Hey and Trefethen, 2003) that need to be archived and preserved so that they are available and appropriate for contemporary discovery and future re-use (Lord et al., 2004). Policy level reports in the UK (Lord and Macdonald, 2003) and in the US (National Science Board, 2005) provide national overviews of data curation but also more generally applicable observations of the many kinds of digital collections requiring preservation and the wide spectrum of activities and challenges involved. They describe the current inconsistent state and status of the field, where provisions for data are more advanced in basic life sciences and "big" collaborative science programs. They also reveal that the definition of data curation as well as what it comprises have not reached consensual formulations. However, the reports have started to direct attention to data curation at a time when research communities at large are only beginning to build understandings of and capacities for handling data stewardship issues.

While there has recently been an upsurge in data management "lessons learned" types of papers that stay true to the technical orientation of e-Science, the actual work and full scope of data stewardship remains unexplored. This paper presents one such account through an ethnographic study of, and the authentic voices from, an ongoing ecology collaboration that represents one of the longest-running efforts at long-term consistent data collection, namely the Long Term Ecological Research (LTER) program. LTER represents a particular kind of setting for data stewardship, characterized and challenged by a long-term science perspective coupled with an open data sharing policy of primary research data in a highly distributed environment of interdisciplinary collaboration. Similar to what has been found about the diversity of actual practices in relation to data access (e.g. Hilgartner, 1995) and data sharing (e.g. Van House et al., 1998; Birnholtz and Bietz, 2003; Zimmerman, 2003) in different science settings, we suggest that a variety of data curation and stewardship practices and approaches exist. Through the LTER account in which scientists' quotes are marked with "S" and information managers' with "IM", we hope to enrich the emerging understandings of data curation and to raise some fundamental questions relating to the differences between the LTER data stewardship approach and the more general formulations of data curation in e-Science.

In Section 2 we describe the context and challenges of the LTER approach to data stewardship. We first explain the LTER network science imperatives and then focus on the challenges involved in data sharing and stewardship. While in favor of open data sharing, the LTER scientists' experiences with data sharing are influenced by what is in practice a complex social process in which scientists have to balance different pressures and interests (Arzberger et al., 2003). LTER information managers, in turn, are motivated in securing long-term data (sets) and face the challenges posed by the nature of long-term ecological data and the complexly relational nature of data in their scientific, organizational and technological environments. There are a dearth of studies on actual practices of data curation and stewardship so we frequently relate our findings with policy level reports on e-Science, cyberinfrastructure and data curation. We aim not to provide a fully developed typology of data curation features, rather we identify what we see as the most salient characteristics contributing to the way data stewardship is handled in LTER.

Section 3 describes in more detail some of the practices LTER information managers employ to support the LTER network science mission. The analysis has been inspired and informed by ethnographic studies of work (e.g. Suchman, 1995) as well as the notions of infrastructure initiated by Star and Ruhleder (1996) and communities of practice by Lave and Wenger (1991). We describe the practices through which information managers attend to the extended temporal scale of long-term research and data sets, maintaining continuity by attending both to the histories and to all the ongoing change processes in the scientific, organizational and technological environments at local and network levels. The LTER information managers' experience with data stewardship suggests care is needed both for the record and also for the development of information infrastructure. Our analysis has been informed by studies of technical support work and by research on so-called alternative design approaches that create room for non-professionals in systems development and recognize the various kinds of integrations and intermediations of design and use (e.g. Henderson and Kyng, 1991; Christiansen, 1997; Dittrich et al., 2002; Sandusky, 2003; Kanstrup, 2005; Pipek, 2005).

In the discussion we take up aspects of the long-term and continuity issues by relating them with CSCW and Participatory Design (PD) literatures. We discuss the possibility of enriching the prevalent technology emphasis in e-Science through CSCW types of studies of actual data stewardship practices in a manner more attendant to the inter-related and continuously changing nature of technology, data care and science conduct.

We raise the issue of balancing as a necessary method in managing some of the unavoidably contradictory aspects of long-term data care and infrastructure work. We discuss the competency involved in data stewardship that combines long-term data managing and information infrastructuring. Finally, we make some observations with regard to the recent conceptual developments within the CSCW field pertinent to data curation and stewardship.

We conclude by proposing that more studies of the long-term perspective and the actual data curation and stewardship practices and their full environments are urgently needed to balance the technical overemphasis inherent in near-term planning and with increased computing power, middleware, and shared grid capabilities. Our goal is to consider bringing closer together the visions of e-Science and understandings of actual practices.

2. Data sharing and stewardship in LTER

2.1. LTER NETWORK SCIENCE DRIVERS

The Long Term Ecological Research (LTER) program was initiated by the National Science Foundation (NSF) in 1980 to augment the more typical ecological studies defined by short-term timeframes. While ecology traditionally is concerned with observing the inter-relationship of organisms with their biotic and abiotic habitats over short periods of time, long-term research in ecology is essential for revealing and understanding protracted phenomena, such as slow processes or transients, episodic or infrequent events, and processes with major time lags (e.g. Likens, 1989). Historical change is a key to understanding the present and anticipating the future as well as to the formulation and testing of ecological theory (Callahan, 1984; Magnuson, 1990).

US LTER sites – the number has grown from an initial six to the current 26 – design their own field measurement programs, choosing a locally relevant research focus relating to the five LTER core research areas (Hobbie et al., 2003). Sites conduct field, laboratory and theoretical investigations on their particular biome where biome's range from cold polar to hot desert regions and from tropical rainforest to watersheds and marine systems. Each site addresses patterns and processes that extend over local to global spatial scales and operate on annual to decade to century time scales. For instance, there are experiments covering a 200 point geo referenced survey of an urban-desert site (Grimm and Redman, 2004) and temporal sampling plans for a 200 year experiment on the decomposition of tree logs (Harmon et al., 1999).

Collaborative site science within the LTER with its focus on biomes includes multiple disciplines and themes from soil chemistry to stream flows to forest ecology, each a research and management area in its own right. Such local work has been preparing LTER participants for the efforts of federating across site science themes: while interdisciplinary research proceeds at each site independently, participants also join together for cross-site work. The LTER network is organizationally federated through annual scientific meetings, periodic "All Scientists" meetings, cross-community committees, and shared experiences from working groups to review panels. In addition, the growing attention to informatics, education, and social sciences initiates an interdisciplinary coordination within which jointly framed questions create new types of data needs and an arena within which data integration can be explored. In working together under an overarching theme of the local biome and the biome within its larger social arena, data collected for original empirical inquiries may be pertinent to subsequent research topics that bring together an unplanned group of related data sets.

The "nature" of long-term research provides new understandings of the collection of data over regular longitudinal periods, traditionally categorized as monitoring. Working with such data, collected at a variety of scales and from a variety of disciplines, builds from the concept of monitoring and adds vital temporal, spatial, and interdisciplinary data-intense perspective to LTER biome research. Such long-term studies have brought about shifts in understanding of research questions and have opened up scientific discussions. For instance, a study of storms and their consequent ecological disturbance in coastal regions (Hayden, 2000) contributed to a network theme of disturbance not as an anomaly but as a natural part of an ecosystem's development; an investigation into the relationship between primary productivity and species richness (Waide et al., 1999) brought with it an early understanding of data integration challenges arising from differing methods and heterogeneous data. LTER is accumulating experience interpreting long-term data sets (Hobbie et al., 2003), examples include capturing the onset and causes of acidity in rainfall (Driscoll et al., 2001) and the revelation of a decrease in Northern Hemisphere ice duration (Magnuson et al., 2000). From the information management point of view, long-term science is concerned with the research need to collect and keep records of the same measurements over long periods of time. Long-term studies require data to be recorded consistently, documented adequately, archived digitally, and accessible electronically. Instruments, technology, and analysis methods may change but the emphasis is on maintaining the coherence and continuity of the record of observations over time. Long-term datasets serve as points of engagement for multiple investigators and as entry points for both research and data integration.

Global issues, such as biodiversity, global climate change and ecological sustainability, encourage scaling up of research beyond the US LTER to international collaborations and to increasingly interdisciplinary undertakings to study complex issues of great human concern. Whereas much of traditional

Table I. LTER Network science drivers

Type	Scope	LTER science driver
Science	Biome	Site ecosystem research
Science	Temporal	Long-term research e.g. decadal patterns
Science	Spatial	Global research e.g. global climate change and ecological sustainability
Science	Social	Collaborative research e.g. network of sites and multi/interdisciplinary
Culture	Social	Publicly funded research e.g. mandate for open data access

ecology has been characterized by single investigator studies with strong ties to data and while ecologists in general have experience with sharing data between close associates, research directed toward environmental problem solving is one area where data sharing has been strongly encouraged (Gross et al., 1995; Zimmerman, 2003). Within this setting, data stewardship has influenced the LTER network approach. It has provided a community reminder to associated researchers about the value of continuing an established measurement series rather than abandoning them in order to support new experiments. The value of local data stewardship is that data management understanding, engagement, and planning for change are developing along side the scientific understandings, so that data enriches directly and immediately scientific research and research integration, while site and network science frames the focus of data management activities. Table I provides an overview of LTER Network science drivers and their scope as discussed above.

2.2. CHALLENGES OF DATA SHARING

Sites are selected to become part of the LTER program through a competition held by NSF. After the initial competition, they no longer compete against one another for continuation. Rather, the intellectual integrity and coherence of a site's development is reviewed midway through each funding cycle and judged at the end of 6 year funding cycle through the assessment of renewal proposals by a panel whose criteria include scientific progress as well as the degree of cooperation. Evaluation addresses also a site's concordance with the NSF mandated open access policies. Since the mid 1990s, NSF has aligned LTER funding with open access policies directed toward publicly funded scientific and technical data (cf. Arzberger et al., 2003) and has required each site to have primary research data available on the Internet 2 years after its collection.

The requirement to make data accessible is argued on the grounds of public funding also within LTER: "I like the philosophy of the LTER, that we get to do our research at the tax payers dollars and the data ultimately are owned by the people'' (S). However, the accessing and sharing of data is also a complex social process in which researchers have to balance different pressures and interests (ibid.). ''I still think scientifically credit should be given to those who originally gathered the data'' (S). As experience in LTER is gained in making data publicly accessible, scientists' hesitation related to data sharing is gradually shifting to a broader understanding of the ramifications of putting data online:

> ''There was some thought of 'somebody is going to steal my data'. But as we thought about it more, as we gained experience, we found there is much more to be gained by making your data publicly accessible than there is by holding it back. [...] Although I think that there are still people that are a little bit hesitant, I don't think it is nearly as controversial as some of us thought 5 or 8 years ago.'' (S)

Improving access to and sharing of publicly funded research data is an issue that touches on all aspects of the research enterprise and the development of knowledge (ibid.). Thinking about ecological research solely in terms of short-term projects or individual scientist's careers has been changing as scientists are challenged to share their expertise in collaborative forums and perspectives are broadened in terms of more synthetic and longer-term planning of the research projects:

> ''There's an element of a social definition of us LTER scientists: someone who is willing to maybe not do all their research in this narrow way but to do some of it in a true collaborative manner so that something new comes out of it that wouldn't come out individually.'' (S)

Collecting and preserving long-term data for creating a data legacy is an important element of the LTER participants' responsibilities. Similarly with emphases on scientists' responsibilities as data creators (Lord and MacDonald, 2003; National Science Board, 2005), LTER scientists are encouraged to plan and prepare data for future use by providing context for their data and documenting with metadata. For the individual researcher, the sharing of data particularly prior to publication can be burdensome, time consuming and unrewarding (Arzberger et al., 2003). The work of documenting data is highly problematic from the point of view of scientists:

> ''One has to be productive, but we also have to make some investments in the long-term. Scientists in general are sympathetic to getting a publication out. With data documentation they realize that it will probably take them 20 or 30% more time if they actually really clean the data up, figure out what it is and get it stored away properly. And some people 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.'' (S)

The need to motivate researchers in their new data challenges by appropriate reward structures and to change the culture of the research environment to be more supportive of data creators in conjunction with curation activities has been recognized (e.g. Lord and MacDonald, 2003; Arzberger et al., 2003). However, currently it continues to be a contribution to community for an LTER scientist to take the time to make their data usable by others, particularly when there is not necessarily any immediate return anticipated on their investment in terms of scientific merit or reusable data from elsewhere. This is also reflected in scientists' views of information management; these views have changed over the years but there still remain a variety of expectations and priorities.

> "Information managers are expected to take all the messy data and get clean and make it available to scientists. (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 scientists 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." (S)

2.3. CHALLENGES OF DATA STEWARDSHIP

LTER information managers share motivation for supporting long-term data as they are aware that: "A database increases in value over time if well maintained, even though it may lose some of the historic facts, the overall value will increase" (IM) and of the inevitable data decay if data are not well curated: "Loss of information can happen remarkably fast. We have had anecdotes about little studies that were not properly documented. When they attempted to go and recreate them 2 years later, no one could remember how they had been done." (IM)

LTER sites collect largely observational data (LTER data activity characteristics are summarized in Table II) that contribute to an understanding of the local ecosystem and to central program themes. Some sites have designated "core" monitoring datasets as those that will be measured and maintained over extended periods of time, whereas some other sites have decided to preserve all data collected on their site's premises (cf. promotions to archive and curate extensively if not all-inclusively observational data because it is hard to reproduce or non-reproducible, e.g. Lord and Macdonald, 2003; National Science Board, 2005).

Ecological research typically deals with heterogeneous data, that is small and diverse, non-standard datasets are most common to ecology:

> "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. In genetics, for example, in comparison, databases are deep but not as complex." (IM)

It is the variability of ecological data that makes them particularly difficult to describe adequately enough for others to use (e.g. Zimmerman, 2003). Furthermore, ecological data sets are often extremely complex because: "missing values, midcourse modification of sampling or laboratory procedures, addition or deletion of study parameters, personnel turnover, plot or habitat modification by disturbances (natural or anthropogenic) or changing environmental conditions, and numerous other factors leading to data anomalies are commonplace" (Michener et al., 1997, p. 332). The data taking itself may differ and require documentation regarding collection methodologies in practice or new techniques deployed in the field. Such research data require extensive quality assurance and control before preserving them in a public database.

Due to the long-term perspective of research, LTER data sets are under continual change. Typical examples of "dynamic data sets" both accrue annual additions and are subject to various kinds of revisions, whereas "static data sets" once collected are closed and need no further curation (Lord and Macdonald, 2003, p. 52). The revisions LTER data sets face range from "methods changing to the questions asked of datasets changing over time" (IM), even to "the thoughts on why it's being collected and should it continue to be collected changing" (IM). The dynamic, ever evolving type of datasets typical to LTER stresses the need for continuous data management: "It's a never ending battle to really keep the information preserved over time, especially when you are still collecting it" (IM). This is elaborated in Section 3.1.

Data collecting in environmental field science have always involved the logistics of operating within an "outdoor laboratory". Coupled with this, ecology has a history of manual data taking. The traditional hand written field notebook or group station log of activities creates flexibility in practice with the possibility of in-the-field category modifications or inserted margin notes. These unexpected reorganizations and notations represent science-in-the-making (e.g. a margin note may be an alert about a previously unnoticed shadowing of an incubator or about an unplanned change in the equipment), yet create challenges for structured data flows and present challenges to update while data collection continues. Extending electronic techniques into the field, e.g. through use of digital notebooks, lessens post-field data ingestion difficulties but requires designing for emergent field understandings as well as for standardized input. New experimental techniques have started to enter the field, changing the conduct of field science with automated data collection producing new types and large volumes of data. Thus they require a significant amount of planning to reorganize for new data collection procedures. These may augment or update existing practices, arrangements, and resources; they frequently require cross-calibration with previous techniques and subsequent analysis together with existing data collections.

Traditional scientific data sharing has occurred though journals with publication of data that has been checked, analyzed, and considered within the context of experience and other data. In this process primary data is transformed into derived data and synthesized data products (Lord and Macdonald, 2003; Bowker, 2000). The LTER community decision to make primary digital data available 2 years after collection introduced a new focus on primary data posted online rather than on secondary and tertiary data or summaries available in articles. This represents a significant augmentation of responsibilities moving from the need to understand and synthesize materials within scientists' career and project timeframes to requirements for contextualization and preservation of primary data for re-use over longer-term timeframes. LTER sites are faced with defining the work and data preparation to transform previously tacit, informal understandings regarding data nomenclature, methods, context, and quality into explicit procedures incorporated into updated practices.

In the LTER network the expectations for data access have evolved from well-curated data for site science to a focus on intersite availability, and to open public access.

> "When LTER information management started, the idea was to share data within the sites; focus was on site science, site data management, even that was a revolutionary idea. People knowing one another, trusting one another but more difficult inter-site-wise. Since the emergence of the Gopher technology and World Wide Web, there has been the demand to put data online. This was a new expectation, not just that data was managed well and documented but to be made readily available to wider community." (IM)

Open access to research data requires an extensive contribution to providing context to data to be truly accessible for public use, and more standardized ways of describing data are needed as is elaborated in Sections 3.2. and 3.5., respectively.

The nature of long-term, collaborative science provides various kinds of uses and reuses for the collected and curated data. Long-term data defies the simplistic definition of "reuse" or "secondary use" as it entails "the use of data collected for one purpose to study a new problem'' (Zimmerman, 2003, p. 7). During its life-cycle, an individual long-term data set can have relationships with other data sets and research questions that can be very complex. First, there is the ''monitoring'' aspect of data in which records are added annually to a dataset. Short-term data use yields a gradually developing, more informed understanding of the local ecosystem. Analysis of annual additions in association with other long-term data sets may lead to new hypotheses requiring more data gathering or to immediately publishable results. An individual LTER site provides a collaborative prompt to associated investigators to think outside their individual research topics and to integrate across investigator studies given overarching shared themes such as the local biome. The consequent sharing of data is indeed a new use of the data. In addition, the very planning of joint experiments or of synthesis work is where tacit, informal, and formal data definitions, clarifications, and relations first develop. Investigators within the LTER network are exposed and encouraged to develop and engage with cross-site themes. Such arrangements often encompass data considerations as well:

> ''The encouragement to do cross-site synthetic work has been there for a long time. A lot of the discussion in the information managers' group is how to facilitate that electronically. First, how do you make these resources available on-line, then how do you make them available in such a way that they can support cross-site research.'' (IM)

Data integration is an important feature of data referring to the ability to align and/or combine data. As part of the LTER ''third decade synthesis'' endeavors, community efforts are focusing on data integration and data exchange. Such efforts contribute to the ability to federate or centralize data. Both within and outside LTER, data modelers are supported when long-term data are available. Downloads of LTER data for reuse by non-LTER scientists and the public are another type of data reuse.

Table II. LTER data activity characteristics

Activity	LTER characteristics
Data taking	Site specific ecological and social science data Observational, largely non-reproducible data Heterogeneous and complex data (sets)
Data preserving	Dynamic data sets, annually/seasonal updates
Data describing	Multi-site data category building
Data using	Long-term site and network science
Data sharing	Open, public access to data and metadata 2 years after collection
Data reusing	Appropriate data structure, context and presentation for interoperability All reuses cannot be anticipated

3. Problematics and actual practices of data stewardship in LTER

The kinds of solutions for data stewardship that LTER information managers have developed have been influenced by a range of science drivers and data activities described in Section 2 and summarized in Tables I and II. In what is known as a cooperative, federated database system approach to organizing information management in LTER (Baker et al., 2000), the site level data stewardship consists of ongoing, retrospective–prospective data management, intensive data contextualization and description, as well as judicious technology design. At the same time, on the network level, information managers engage in collaborative information infrastructure and metadata standards work.

3.1. ONGOING DATA MANAGING WITH EXTENDED TEMPORAL HORIZON

LTER information managers attend to the various challenges posed by the LTER data and science environment through ongoing data care activities that range from retrospective rescue to prospective planning. The extended temporal horizon of data managing depicted in Table III showcases past, present, and future through quotes about how information managers recover legacy data, attend to the ongoing maintenance of core datasets, as well as anticipate and prepare for the future.

Perhaps the most obvious activity along the temporal horizon of the LTER data management is attending to the ongoing data collection. Although sites have procedures and protocols for this repeating cycle of data collection and

Table III. The extended temporal horizon of ongoing data managing in LTER

Recovering legacy datasets	Attending to ongoing data collection	Designing for the future
"I was trying to document a lot of historic stuff because the PI [principal investigator] was coming on with Alzheimer's and I knew that he was going to retire. I had a series of interviews with him and I got incredible documentation for these early corporate data." (IM)	"Getting scientists' data into our system from the very beginning...whether it is to help them with data entry forms, setting up data entry programs, all the way from QA/QC programs to getting it archived into our system and accessible on the internet." (IM)	"We envision 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." (IM)
Historical/legacy	Immediate/near-term	Long-term

archive activity, ongoing data collection nevertheless requires both care and service work on the part of information managers (cf. Star and Strauss, 1999). This work is described in the center column quote in Table III by an information manager. New data are taken every season of field and laboratory data gathering. Quality analysis and control are performed together with updates to metadata. These elements of collecting, cleaning and using the data are part of the traditional recurrent cycles of short-term data use and publication. The long-term perspective further necessitates careful aligning of any new data accrued, assuring it "fits" with and continues an existing collection. This typically requires meticulous documentation of changes that have occurred. Superimposed upon this is the NSF prompted and LTER enacted 2 year data policy that requires scientists to submit their data and metadata to a local repository. This requirement represents a new view of data: it includes a responsibility to the site in terms of capacity for synthetic research, to the future researchers who may ask new questions of the data, and to the public who may use the data to inform and to frame environmental policy actions.

The first quote in Table III gives an example of data management activities carried out retrospectively. Here an information manager voices the excitement of having been able to recover some valuable data sets and record additional historical context information. Efforts such as this one are precious, not only to prevent the loss of a site's particular set of continued longitudinal studies and development of a long-term perspective, but also for global ecological research (cf. the work of identifying valuable ecological data sets at risk, Gross et al., 1995). Recovering the "past" requires extra devotion because the potential value of a data set in the future is very difficult to evaluate (Lord and Macdonald, 2003; National Science Board, 2005). Continuous prioritizing of what gets done with the resources available has to be carried out as the tendency is for more acute matters to take precedence.

The prospective dimension of data management activities is captured by the last quote in Table III where an information manager explains the need for metadata standards as one element of designing data infrastructures for the future. On one hand, she voices commitment to implementing standards for the benefit of the collaborative effort. On the other hand, she carefully considers how they should proceed at their site because of both the expected highly labor-intensive undertaking and the unexplored problem area. Designing for the future is complex: to address the problematic issues inherent in data management, a balancing is needed of (meta)data and technology issues and comprising activities both at local and network level (Sections 3.3.–3.5. elaborate on this).

There are predictable elements with regard to an extended temporal horizon of ecological data management that include the immediate-term issues of seasonal and annual cycles of data collection, entry to databases and preservation together with getting the metadata; the near-term issues of data use and publication resulting from the 2 year data policy stimulus for scientists to submit their data and metadata; and the long-term issues of data reuse and synthesis. The retrospective issues of recovery of valuable datasets are less predictable but may require rather urgent attention. When and how to migrate elements of an information environment remains uncharted territory, and often impossible to link with established or fixed time scales. Developing the juggling skill to bring together the various time scales into a working whole involves a great deal of tacit knowledge for prioritization and is an assumed part of an information manager's expertise:

> "When I first started my job, I found it very difficult that there would always be some things that I thought needed to be done but I could never get to. I kept having to do triage everyday 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, and occasionally to require more resources." (IM)

3.2. INTENSIVE DATA DESCRIPTION

Creating a legacy of well-designed and documented long-term experiments and observations for use by future generations requires scientific data is accompanied by contextual information that describes the data collections. These descriptions are called metadata (data about data). There are two dominant reasons for LTER intensive data description requirements. First, small and highly diverse ecological data sets themselves present particular challenges in terms of documentation (see Section 2.3.). Such variability necessitates that contextual information for ecological data are carefully recorded starting with the data planning and subsequently with the actual data taking and data curation. This is best carried out at the site level:

> "An unwritten rule is that each site manages their own data. Data are best managed at a site by people who know them. That may or may not include archival, as datasets grow and become unwieldy, to be able to store and protect them. But their management, as far as quality control and assurance, and understanding the ways in which they were collected and the sites that collected them. There is a real feeling at sites that the best place for data is at sites."(IM)

Second, in the context of long-term collaborative research and open access to data, the many ways in which data may be used and reused need to be accounted for in the data descriptions.

> "You have certain levels of metadata that you would describe. If someone within the site was using it, they know a lot about the whole collection system and the research system at the site. You can give them less metadata, just specific data about the actual dataset itself, you wouldn't necessarily have to give them the broader view of the abstract and everything. It is not as critical, but to somebody outside or for somebody 20 or 30 years down the road, then it's going to be more and more critical that this whole story unfold." (IM)

The quote reveals the complex situation for data description brought about by the temporal and spatial/geographical, institutional and interdisciplinary distribution of the reuse situations, as well as by the different, at least partially unknown, future user groups and their needs (cf. Markus, 2001). The information managers are crucially aware of their role and responsibility in making choices with respect to data stewardship and metadata description on behalf of the current and future user communities (this is called community-proxy function by National Science Board, 2005).

Data description is an essential but also controversial issue in LTER because the process of assigning metadata and otherwise documenting small, observational data sets is highly time-consuming and labor-intensive (Gross et al., 1995). Often overlooked is that methods for data description are the subject of ongoing research (e.g. Birnholtz and Bietz, 2003; Chin and Lansing, 2004). A rather consensually held view about metadata provision in LTER (also in e.g. Lord and Macdonald, 2003; National Science Board, 2005) is that the data originators, i.e. the people who know most about the data, its nature and the actual data taking, should also provide the metadata. However, this is not yet always the prevalent cultural norm and actual practice. In LTER information managers are responsible for data management, including the metadata issues: "We really need to promote mechanisms for being able to get the data as well as getting the descriptive data, and the metadata is really the most difficult" (IM) and "you need to put a carrot on a stick and to follow-up on that" (IM). This work includes raising awareness of data, and educating investigators and students about the importance of data management:

> "You need to convert them into thinking that putting 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. [...] It's been a massive process of sort of education and badgering, to get people to think that that is important." (IM)

Information managers provide various mechanisms for metadata provision: "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" (IM) and "The easier you can make it for them to put the metadata in, the more success you can have" (IM). They oversee: "It's a matter of leading them through the process" (IM), and make available various kinds of data management services: "We offer them whatever systems and help that will help them do that" (IM) as well as incentives: "You can develop the information system to do quality assurance for them" (IM). And if these do not work, information managers have developed work-arounds for gathering metadata, for instance, one tells: "A lot of the documentation I end up pulling out of the paper, methods and things that they are not going to give me directly. I end up tapping other resources as much as I can." (IM)

Furthermore, there are instances in which information managers are the best people at their sites to provide metadata for datasets. For instance, some sites have data managers work part time as field technicians collecting the "corporate" or "core" data, who, hence, are able to develop both an intimate, local understanding of the data together with a deep knowledge and concern with the handling of data. In other cases, long-term involvement with particular data sets provides privileged knowledge about and appreciation for the data:

> "I have been there for a long time, and have developed all those systems that deal with the climate data and particularly the stream system and stream chemistry [...] we have had these stream flow data, and our climate data, and some of the survey data that we do from 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. These have been going on for so long that none of the PIs are the ones that originally started them. A PI may get assigned to the dataset[...]but he doesn't understand how it is processed or anything. So I end up doing all of the maintenance, all of the metadata, I mean that dataset wouldn't be a nice long-term dataset if it wasn't for the data manager". (IM)

Securing a "nice" long-term dataset by maintaining the data and writing the metadata shows a profound understanding of the data, their ecological context and connection with the studied domain as well as collection for long-term research purposes. The discussion of domain scientists and the information manager illustrates how well positioned the latter is to understand both the data with their connections to an actual domain and research, and to the technology used in processing and archiving data. (Karasti and Syrjänen, 2004.)

3.3. JUDICIOUS, SITE-LEVEL TECHNOLOGY ALIGNMENTS

Information managers play active roles in how technology and data management concepts are introduced and sustained at sites: "Researchers are looking at the information manager for guidance. Information managers need to be proactive and come up with their vision and plan for the site, where the site is going and how it connects to the network information management." (IM) Information managers, responsible for adapting and aligning new technologies and data management at their site, have to pay special attention to existing site practices and technologies as well as available funding possibilities.

As technologies are developed at increasing speeds, staying technologically informed is an important aspect of an information managers' work: "It is a constant battle to keep up with things, to remain current in technology" (IM). Although staying technologically current is a major driver, other factors that relate to the long-term perspective underscore the merits of modest and unadventurous approaches in site information management systems. Information managers need to be able to understand the local research in technology considerations because providing support for site science is the initial, immediate, perennial responsibility of information management: "I see it's important that information management is driven by the research, that information managers continue to come back to assessing whatever projects they want to develop to whether it is really going to support the research at the site" (IM). As technology is ultimately evaluated against its value for ecological research, information managers are accustomed to thinking about and designing technologies embedded within their social and organizational contexts of scientific work and collaboration. Local approaches may, therefore, have legitimate differences in their emphasis and methods of technology development: some sites prefer to "keep it simple" (IM), some emphasize "data availability, data accessibility and possibilities for exploration" (IM) and some go after "automating systems and experimenting with new technologies" (IM).

On one hand, incorporation of new capabilities to enhance data capture, use and preservation always holds the potential for extra facilitation of science. On the other hand, there is the concern for having in place a data-safe functional system, "a protecting cocoon" (IM), for maintaining the integrity and availability of the long-term datasets. Judicious decisions about technology procurement are influenced by the features of high reliability, easy maintainability, and low risk for long-term data management. An information manager's foremost concern in aligning developing technologies with existing technologies and practices is to minimize disturbance of ongoing data archival and use followed by interest in optimizing long-term data re-use: "The experience we have had with several of our things [...] 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).

In 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). The evaluation process that places research sites under scrutiny periodically sets a timeframe for some technological updates: "We manage to update web pages every 3 years, for review and proposal. We are on this cycle, and we end up putting a lot of energy into updating." (IM) However, transitions of a larger magnitude occur less often, and the persistence of technological change prompts cautious thinking and careful balancing of options: "Having the investment in [current technology], it is not so bad yet that I would want to go and rewrite all my interfaces" (IM) and "We are transitioning our whole design, we are really facing a lot ... then it stabilizes again. Every so often things need to migrate, the technology changes so much." (IM) Alignment becomes increasingly complicated as legacies grow because interdependencies between data, technologies, organizational, and institutional practices grow. 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). Infrastructures are embedded into and inside other structures, social arrangements and technologies (Star and Ruhleder, 1996).

Within the LTER network information managers are often seen as the proponents of technology despite their rather unadventurous, "feet on the ground" technology approaches: "The information management community in LTER has been extremely proactive, and very responsive to demands not only at the site but, in fact, extraordinarily responsive at the network level" (S). Information managers themselves see that their proactive role with technological infrastructure is "really pivotal in leading the community in recognizing the value of information technology and information management" (IM). Some of this tension relates to the necessarily different points of view of the role and phase of technological development.

3.4. COLLABORATIVE INFORMATION INFRASTRUCTURE WORK

Anchored by the realities and needs of their sites, information managers have created a network level forum, the LTER information managers' committee that forms their "community of practice" (Lave and Wenger, 1991). Information managers bring with them to their network level activities comprehension for local settings and appreciation for the diversity of local infrastructures: "One critical thing to the success of the information managers' group has been the recognition that there are legitimate reasons for some differences between site systems. [...] In terms of the IM group, you accept also views different from your own" (IM) and "There is a variety of approaches among sites, and there is strength in diversity" (IM). In addition, information managers share interests in technology and data issues that cross geographical and ecosystem bounds: "The network is not that cohesive as far as science goes, every site is independent. [...] It is really the information managers that have created a network framework because there is a sort of a synergy of interests for us." (IM) Information managers are united by "really believing in what we do" (IM) for long-term data in long-term science context and pride for the group: "We are an incredible asset to the whole LTER program" (IM). Furthermore, the network level venue provides information managers a special point of view to gain an understanding of the LTER network level activities and an arena where information manager's voices can develop: "We have made a greater impact as a group" (IM).

Awareness of the long-term provides an opportunity to develop a community with continuity:

> "The long-term has the advantage that you know that you are going to come back to things. If a thread slows down or is dropped, down the road you can pick up that thread. You will readdress something the next day, week or year. You are always related, affiliated, associated. LTER has that continuity." (IM)

Continuity creates the confidence within the community needed to be able to interact regularly: "not having to rebuild every time, we have created trust" (IM) and maintain reciprocity: "It is good to see how other sites are doing things, either as a contrast or as an idea to improve" (IM). It provides a reliable place for sharing: "It is safe to say things that demonstrate examples of where people have not been as successful, or disappointments. As soon as you are able to do that in a group, there is a bonding that occurs" (IM). Information managers gain knowledge together: "LTER information managers have taken the time that fosters an integrative, sustainable approach with technology, ensuring that we learn together. [...] It's all like being mentored by the overall group" (IM).

This network level community offers an arena for collaborative information infrastructure work. Information managers have created approaches for jointly designing shared infrastructures that rely upon on the inherent characteristics of the networked organization (see also Karasti and Baker, 2004). One of them draws on the technological diversity engendered by the network sites that can be used for sharing ideas and learning from each other's experiences. One information manager describes this federated way of using local heterogeneity to the advantage of the network as a "cherry-picking octopus":

> "One of the advantages with 26 sites is that there is always someone doing a major upgrading: they're out there looking for the solution that would work the best, they might find the solution through information managers' meetings, word of mouth, DataBits [LTER information

> managers' network newsletter], and it may also solve my problem (Boom!) I spend a fair amount of time looking around what is going on within the network. I learn a lot by looking at other LTER sites. If I see they are doing something neat, I'll try to find out how they did it. Cherries are the good pieces of software and there are 26 opportunities to find good ways. It needs to be an octopus as they need to be connected." (IM)

This quote illustrates how the information managers' community welcomes and is willing to consider all potential discoveries of technologies suitable for the ecological research domain. It also demonstrates how the technological heterogeneity at site level is not only allowed for, but that it has been turned into a common resource of proved technology experiences through courses of network wide selection processes where each site is a "laboratory" with its local specificities. And importantly, it showcases a principle that is widely recognized within the LTER: "You can have a lot of the good ideas come from the bottom and work out, not top-down" (IM).

In LTER information management "a lot of the bottom-up characteristics are important" and therefore it has become of utmost importance to be able "to deal with heterogeneity not by limiting it but by dealing with it" (IM). The following two examples illustrate how information managers develop information infrastructure through a common struggle with diversity and consensus, how they take into consideration the local level and account for the legitimate differences between sites in the development of technologies as well as network solutions and standards. First, the LTER tradition of "prototyping into consensus" is based on the idea of the snowball model of incremental participation. In practice it means that each module effort, e.g. a queriable all-site climate database or the conceptual design of a Network Information System, is led by an interested information manager who coordinates design, presentation, and communications with the LTER community throughout development and implementation. Interested sites are frequently recruited to serve as test users so the module becomes a "boundary object" (Star and Griesemer, 1989) that is shared and discussed, redesigned and modified. Although only a few sites may participate originally, discussions during presentations or break out groups at annual meetings elicit the voices of the larger community in which all participants as decision-makers directly influence design.

The second example illustrates a typical LTER information managers' group approach to developing guidelines and consent approaches that outline a minimum set of requirements. It draws in the larger LTER community for decision-making and agreement forming on a policy that has been jointly developed over time and engenders flexibility and openness to accommodate the variety of sites.

> "In IM meetings we brainstormed some basic principles for information management policies. We took that to the CC [LTER Coordinating Committee] and they appointed an ad-hoc committee, which in some cases consisted of people who were most recalcitrant about these things. They made some changes, even strengthening some points what we had suggested. The basic principles were rather 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. [...] We did not come up with the LTER wide information policy; we would have ended up in endless discussions. We published guidelines for individual site information management." (IM)

LTER information manager's collaborative information infrastructure work, as illustrated by the above examples, shows an enduring, tentative and open relationship between infrastructure and the conventions of a community of practice (cf. Star and Bowker, 2002). Recently, the adoption of the concept of the ecological metadata language (EML) standard has both prompted outside the LTER network collaborations and intensified the challenges for being able to account for diversity and flexibility in the processes of metadata standards work.

3.5. METADATA STANDARDS WORK

The need for data documentation is accepted in LTER as we have described above due to the nature of ecological data, the way it is collected and the extremely complex (re)use situations (see Section 2.3.). Metadata has a crucial role in this: "Metadata is really unlike anything that has been done in ecology, and it does preserve datasets over time. The network has had a great influence, pushing forward a standardized approach to collecting metadata." (IM) LTER information managers have a long history of exploring data documentation issues: "A major emphasis of the information managers' group since about 1990 has been on documentation. They call it now metadata, we called it then data documentation. [...] The network has had a great influence in pushing forward a standardized approach – not standards – to collecting metadata." (IM) Information managers initially promoted standardized approaches to collecting metadata, not actual standards. They have gained an exceptional understanding of metadata related issues and practices from the point of view of long-term research and maintenance of data and information infrastructure. Information managers are critically aware of limitations with metadata approaches, for instance, about losing the layer of informal description for short-lived, non-scientific narrative data when scientific data are stored with only formal metadata: "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'' (IM) and needing additional informal description:

> ''If people feed us back information about a dataset we put it into the database, then other people can read what they 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.'' (IM)

> ''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, like a 'subjective quality indicator' of some sort.'' (IM)

The information managers recognize that finding more elegant ways to share contextual information is a much harder problem (cf. Birnholtz and Bietz, 2003; Chin and Lansing, 2004).

Currently LTER information managers are engaged in locally enacting the Ecological Metadata Language (EML). They are excited about the new potential though there are site level activities to consider. They see the major information infrastructure redesign: ''You have changing technology, you have expanding metadata content, and that leads to this redesign of this massive schema'' (IM) and recognize the potential pitfalls between comprehensive descriptions and long-term maintenance: ''New metadata standard requires extensions, for instance methodologies to be defined to the level of attributes. This is very difficult to maintain, very comprehensive. I like it but I am worried – it's hard to keep it up to date.'' (IM)

The story unfolding with respect to the EML provides an example of recognizing the need for generic or network standards while taking into account local or site needs and practices. From the outside developer's perspective, the EML standard is successful because it has been endorsed and adopted into practice by the LTER information managers as well as by other groups within the ecological community. From the LTER information manager point of view – an enactor's perspective – EML may be said to be a work in progress since it requires continuing local re-development (Millerand and Bowker, personal communication, 2006). As with other technologies, it requires integration with existing infrastructure. Although mediating tools for EML have been designed and deployed, an array of data issues has emerged where local practices need to be articulated in response to proposed technological solutions. While standards are valuable in establishing a commonality of methods, formats, and semantic content, it is the process of creating standards that is informed by practice and a likely determining factor of success of whether a deployed or adopted standard is enacted in practice (ibid.). Indeed, when viewed as an interoperability strategy (Baker et al., 2005), the standard has successfully engendered dialogue, development, and design while prompting new forms of technical commitment, community involvement and organizational restructuring.

4. Discussion

Issues of long-term and continuity foreground themselves persistently in our analysis of the data stewardship practices and approaches in LTER. Long-term perspective is scarce both in CSCW and e-Science. Temporal and spatial aspects of collaboration and its technological support have been central to CSCW research since its early days, i.e. the variations of same–different and time–place (Johansen, 1988). However, CSCW research and technology development has had marked emphasis on the short-term, on real-time co-located circumstances. With the exception of Kaplan and Seebeck (2001) who discuss the crucial role of time in complex systems, we are not aware of any studies of collaboration over particularly extended time periods in the field of CSCW. Similarly, short-term emphasis is more typical in e-Science. For instance, research on contemporary digital communities that span distances electronically are being explored today through the lens of "collaboratories without walls" (Finholt and Olson, 1997) and scientific "virtual laboratories" (Sonnenwald, 2003); these terms echo an emphasis on the short-term. However, awareness has been forming within the emerging field of data curation in e-Science that more consideration for the long-term or long-lived aspects of data are needed (Lord and Macdonald, 2003; National Science Board, 2005). Such an emphasis approach was a foundational feature in the LTER as evidenced by the use of "long-term" within the program name itself. Table IV juxtaposes short-term and long-term perspectives as a prompt to considering the features and influences of the differing scientific timeframes.

We suggest that e-Science is likely to be confronted with demanding issues of long-term and continuity, particularly as related to data curation issues. One important indication of this is that the estimates on the proportion of dynamic versus static data sets are surprisingly high (Lord and Macdonald, 2003, p. 34); in fact, many data collections and most databases are dynamic. According to Lord and Macdonald the two categories have different management needs so that dynamic data sets that have long active periods also require continuous curation activities (ibid.). An

Table IV. Scientific timeframes and their features

Short-term perspective	Long-term perspective
Technology solution driven	Science inquiry driven
Digital maintainability	Data sustainability
Data deluge concerns	Data sharing concerns
Data grid structures	Information infrastructure arrangements
Metadata enactment	Data description development
Data curation procedures	Data stewardship practices

important observation here is that dynamic data sets and their curation have not achieved adequate attention by research communities, partly due to the current emphases in e-Science Grid and Cyberinfrastructure funding (cf. Buneman et al., 2005).

In addition to the static and dynamic data set categories proposed by Lord and Macdonald (2003), we suggest that another important dimension is the different ways of collecting data, namely manual and automated collection (see Section 2.3.). An illustrative example of their divergence is the different meanings given to the term "data-intensive". In e-Science literature it refers to quantity, that is automated data collection and data deluge (e.g. Newman et al., 2003; Hey and Trefethen, 2003; Lord et al., 2004; Gray et al., 2005), whereas in ecology, it refers to a process of manual data taking of small, heterogeneous data sets or their intense contextualization and analysis. There is little focus on the so-called "pre ingest" activities surrounding data creation in e-Science (Lord and Macdonald, 2003); data collection is subsumed into the data life cycle which is attended to through "comprehensive metadata-enabled scientific workflow systems" (Michener, 2006, p. 5). The objective of these systems is to create environments "whereby most routine data processing steps including data discovery and ingestion, data transformation steps, quality assurance and quality control, as well as many analyses can be largely automated" and each step of the workflow is "accompanied by the automated capture and encoding of relevant metadata" (ibid. pp. 5–6). This implies a mode of data collection and science conduct with associated data curation needs that are profoundly different from the ones prevalent in LTER today.

Manual data taking has flexibility to a certain degree for emergent elements during data gathering and allows for analysis of the data to begin already in the field or laboratory. Furthermore, manual data taking is inherently a question of data collector and their understanding and relationship with the instrument and the ecological site in which data is collected (Karasti and Syrjänen, 2004). Zimmerman found in her study of the reuse of ecological data that the informal knowledge, the "sense" for data, that ecologists acquire as collectors of their own data in the field or laboratory, plays the most important role also in their use of data collected by someone else because it helps them to understand and to assess the data (Zimmerman, 2003, pp. 211–212). Thus changes in the way data is collected are not simply a question of data gathering, but have implications to the complex relations data have with how science has been carried out in practice (cf. Lamb and Davidson, 2005). Several e-Science reports repeat that today's curatorial processes in data intensive fields of research are labor intensive, and put forward the need to evaluate, redesign and automate them (e.g. Hedstrom, 2003). Pushing forward with automated approaches (e.g. OECD Global Science Forum report, 2005), presents a potential danger of marginalizing other approaches. The point we want to raise here is that in fields with traditions of manual data taking, we should study carefully what is at stake in existing practices.

The prevalent view of data curation in e-Science is motivated by the "digital obsolescence" problem, i.e. data in digital form are vulnerable to technological obsolescence and influenced by the Open Archival Information System standard (CCSDS 650.0-B-1, 2002) in that data curation is related to the movement of digital content across multiple generations of technological media (see e.g. Lord and Macdonald, 2003; National Science Board, 2005). Adaptation to technology transitions is formulated in terms of methods, such as migration, emulation and formal descriptions, to preserve content (information) and systems (applications) behaviors over time as successive hardware and software technologies become superseded (Lord and Macdonald, 2003, p. 30). Thus the view of data curation in e-Science is technology driven. LTER network, in turn, sheltered by the exceptionally long and continuous periods of research funding, has been in a privileged position to explore a more science-driven approach to data stewardship that recognizes the complexly relational nature of data in their environments of both scientific, organizational and technological change, and involves socio-technical infrastructure work. Furthermore, the LTER example showcases and pushes us to consider how science, data and infrastructure "grow" together (Fischer and Ostwald, 2002).

In LTER, technology change is intimately intertwined with the changes going on in ecological research that – like all scientific enterprises – continuously reformulates and identifies new questions (Pickett et al., 1999; O'Day et al., 2001). Though change is ongoing, it is not necessarily a simple incremental process, nor a wholesale displacement and transformation (cf. Fischer et al., 2004). First, there is the concern for having in place a data-safe, functional infrastructure for maintaining the integrity and availability of the long-term datasets. Second, the incorporation of new capabilities into the infrastructure for enhancing data's capture, use and preservation holds the potential for synergistic facilitation of science and its changing practices and needs. Third, though reuse of data already takes place in certain forms, for instance, by data modelers and through LTER's third decade synthesis efforts, all of the potential reuses or types of reusers are not known and realistically cannot be anticipated. These are brought together in focused attempts to take advantage of major advances in technological infrastructures that are interposed by interludes of more stability (cf. Hanseth et al., 1996). Similarly, in relation to large evolving systems and information repositories, Fischer and collaborators have proposed "systems that evolve over a sustained time span must continually alternate between periods of activity, unplanned evolution, and periods of deliberate (re)structuring and enhancement." (Fischer et al., 2004, p. 36) Therefore, change in the LTER is informed by enduring, tentative and open interaction between understandings based on the knowledge within the long-term domain of practice, in the experience of using and having developed existing tools, methods and technologies, and in the "leaps of faith inspired by imagination" (Suchman, 2000) in envisioning new technologies. In their work to ensure the longevity and continuity of the network's data and infrastructure, LTER information managers provide support for rapidly developing technology, data requiring continuous "slow time" care, and science having to cope with short-term evaluation cycles of scientific merit and long-term motive (for more see Karasti and Baker, 2004). In balancing between varying time scales and drivers, LTER information managers exemplify how critical it is to integrate these elements in the infrastructure for long-term ecological science in an ongoing manner. In fact, it can be said that to achieve continuity (which is the aim) they need to seriously engage in the continuous work of balancing (which is the method) the different temporal scopes.

Another aspect critical to data stewardship that the LTER information managers have grown accustomed to dealing with through the method of balancing relates to local heterogeneities and global standards, a topic that is well-recognized in design literature that draws from social studies of science and technology (e.g. Star and Ruhleder, 1996; Rolland and Monteiro, 2002; Rolland et al., 2006). The approach information managers have taken prioritizes the provision of the "protecting cocoon" for locally curated data sets, influenced by the long-term science perspective and understanding of the nature of ecological data. Being as stringent and meticulous as practically possible with the very foundation of securing data sets' long-term continuity, acceptance and tolerance for heterogeneities and fragmentation is exercised with regard to other aspects of the information infrastructure, such as local practices with data and choices with technology. In their approach to addressing the tension between local heterogeneities and global standards collaboratively developing common technologies and standards on the network level (and beyond), LTER information managers are closely aligned with views that put forward the need to strike a balance between sensitivity to local contexts and the need to standardize across contexts in relation to infrastructural information systems in large-scale distributed settings of collaboration (Rolland and Monteiro, 2002). A number of studies portray the articulation work needed in balancing the local–global relation from the point of view of the local communities adapting to a global standard or information system (e.g. Star and Ruhleder, 1996, Rolland and Monteiro, 2002, Rolland et al., 2006). In the LTER case, information managers are able to address the divides and boundaries of the local–global tension already in their collaborative activities of developing shared technologies and standards. They have explicitly "designed flexibility" (cf. Rolland and Monteiro, 2002) into the co-constructed socio-technical constituents of information infrastructures, such as the consensually agreed upon sets of best practices or guidelines, the so-called "minimum criterion" that leave room for sites' legitimate reasons for differences in technologies and approaches, and the Ecological Metadata Language (EML) standard and associated software tools. Furthermore, LTER information managers have actually turned the heterogeneous local data and technology experiences into a shared resource and have found ways to work with them collaboratively on more general levels, as exemplified by the courses of network wide processes, such the ones named "cherry picking octopus" and "prototyping into consensus", where each site acts as a laboratory with its local specificities.

LTER information managers continue to evolve various social arrangements in which to account for the fragmentation and multiple divides within the participants' worlds (scientific, technical, organizational, social and natural) as well as the interdependences of ongoing processes. They rely on "multi-voicedness" (e.g. Greenbaum and Kyng, 1991; Schuler and Namioka, 1993) and "partial translations" (Suchman, 2000) as social safeguards for flexibility and openness in their integration efforts of infrastructure processes. Furthermore, information managers themselves provide an essential, flexibility-engendering element that is an integral part of their "generalist" role. As part of their versatile tasks and responsibilities at the site level, information managers become embedded in various ensembles and activities as well as cross boundaries between the traditional divides of use, maintenance and design (cf. Sandusky, 2003). At the same time, on the network level community of practice, information managers have a possibility to develop the expert/specialist aspects of their role at a federated scale in conjunction with the locally grounded work and routine day-to-day practices. The fruitful setting – in essence a learning environment within the LTER network – mixes, though not without struggle, bottom-up, cross-site, beyond the network with top-down cross-fertilization – an advantageous position that gives the information managers the grounded understanding required to integrate the fragmented nature of the ever-evolving and to-be-balanced information infrastructure in a flexible, continuous manner.

The role of LTER information managers has been formed in close relationship with the research elements it supports within the shelter of a long-term science community. Elements of their LTER data stewardship were described in Sections 2 and 3 and are summarized in Table V. The information manager community is characterized by their first-hand understanding of science, data and technology in the LTER context (Karasti and Baker, 2004), and above all long-term vision that pervades – so integrates over and between – all aspects of their work. Their role has been essential in promoting long-term continuity of data, as they have been diligent in data stewardship as well as thoughtful proponents of data sharing and information technologies, understanding at the same time both the local specificities and the lags in immediate results as the scientific needs and conceptual understandings of the data and its use from multiple perspectives mature. In comparison to the "curator" role proposed by Lord and Macdonald which expands the view of curation from archiving and keeping inventory to embracing "the care of the record within scientific context and environment" (Lord and Macdonald, 2003, pp. 44–59), information managers in LTER have even a more comprehensive role as their responsibilities also include standards and information infrastructure work. In fact, their role and variety of responsibilities come closer to the combined categories of "data managers" and "data scientists" put forward by the National Science Board report (2005, pp. 26–27) of which the latter is very broad indeed having emphasis on creativity and intellectual contributions while the former portrays a more traditional data manager position. LTER information managers have versatile backgrounds and often have no clear career path. They typically have formal education in ecology or other science domains rather than in computer science or related fields, thus differing from computer scientists and IT

Table V. LTER data stewardship in practice

Element	In practice
Data heterogeneity	"In general you are struggling with the diversity of different types of data." (IM)
Data quality	"The QA is a big issue, in terms of like curatorship." (IM)
Data description	"... they [scientists] realize that it will probably take 20 or 30% more time." (S)
Open data sharing	"This was a new expectation, not just that data was managed well and documented but to be made readily available." (IM)
Data standards	"The network has had a great influence, pushing forward a standardized approach to collecting metadata." (IM)
Technology sustainability	"... 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)
Information infrastructure	"And they [scientists] simply don't appreciate the time and the energy and the effort required just to do the nuts and bolts maintenance." (S)
Learning environment	"It's been a massive process of sort of education and badgering..." (IM)

specialists. Nevertheless, an important part of their work is technological infrastructure work. Their role bears similarities to those described as appropriators (Pipek, 2005) and local designers (Kanstrup, 2005) with the distinction that all their work is rooted in and shaped by notions of data stewardship. In fact, the pronounced intertwinedness of science–data–technology trajectories may be representative of data stewardship work in general.

Developing information infrastructure often is regarded predominantly as a technical endeavor. This tradition is continued in the case of e-Science where "data grids" are built as data management architectures, specially geared towards satisfying the combined needs of large data sets, geographic distribution of users and resources as well as computationally intensive analysis (Chervenak et al., 1999). In CSCW a substantial portion of studies has focused on singular applications; only a fraction of research has been directed to integrated information systems or infrastructure across heterogeneous contexts. The notion of infrastructure by Star and Ruhleder (1996) recognizes the extended temporal underpinnings involved in long-term ecological research and the extended temporal scope of the ongoing, longitudinal development efforts. Furthermore, it allows analysis to be sensitized towards multi-relational understanding of the socio-technical–historical embeddedness, the transparent and taken-for-granted nature of infrastructures, and the related local–global processes of an installed base and embodiment of conflicting standards.

There are powerful and potentially highly relevant lines of research and conceptual work ongoing in CSCW which, however, would benefit from further development towards the long-term perspective and towards data and technology activities of more perennial nature, such as design-in-use, maintenance and sustainable development, in order to be amenable to ongoing conceptual work in relation to data stewardship. For instance, the potentially promising CSCW concept of organizational memory has recently gone through an overhaul from "only a few uses centered around particular technologies" to further develop towards considering organizational memory both as "object" and "process" (Ackerman and Halverson, 2004). Another example is the concept of common information space (CIS) that, after a period of research on more short-term aspects, has in recent years been explored in terms of the original direction to study "CSCW systems 'on the large'" (Schmidt and Bannon, 1992) and expanded with regard to large-scale collaborations (Bertelsen and Bødker, 2001). Along another line of CIS research drawing on the more recent developments in social studies of science and technology, CIS has been further developed conceptually to deal with the fragmentation and imperfection endogenous to largely distributed contexts (Rolland et al., 2006). There is a related move in the database community to consider "dataspaces" that provide room for coexistent heterogeneity and allow for change (Franklin et al., 2005). Such approaches are remarkable in that they open up and serve well the disintegration and (re)integration aspects of data in heterogeneous settings from the use and management points of view. From a long-term perspective arises the kinds of elements innate to certain aspects of LTER data stewardship work, such as the ongoing performance of infrastructure sustenance concurrent with ensuring the continuity of dataset care. Such a mix of activities and approaches appears common to stewardship work in general, thus analogous to some of the ideas put forward in relation to nursing the chronically ill patients (Strauss, 1975; Strauss et al., 1985) and to considering how architectural memory resides in "how buildings learn" (Brand, 1994).

5. Conclusions

The vision of e-Science includes the potential to both generate and share data through technologies of automated data collection and large-scale resource sharing networks in ways not previously possible in science. While the idea of open access to publicly funded research data is an admirable one, it is also an unresolved concept in practice and poses unprecedented challenges to the actual conduct of science, curation of good quality data, and understanding of long-term stewardship. The challenges involved from the scientists' perspective are not limited to providing proper documentation of the data (which in itself is a contested terrain!) that makes data available and intelligible for more intensive collaborations, the data challenges are profound and pervasive with respect to "doing science" in practice. For instance, an e-Science including only automated data collection would change the very underpinnings of endogenous science-in-the-making kinds of activities common in ecology, and other disciplines with intensive data taking in the field or laboratory. This includes scientists' close relationships with the physical site and the processes of hands-on data taking tightly coupled with the courses of analysis and hypothesis formulation that occur before, during, and after work in the field. Thus the lesson to be learned is to investigate carefully new e-Science technological potentials within their complex environments of science conduct (cf. Jirotka et al., 2005).

Given the multiple work settings and varied uses of data, it is essential to consider thoroughly from the beginning and over time the topic and problematics of data curation. In so doing e-Science would be able to foreground the often invisible work carried out in traditional science practice that has remained backstage in planning of data federation and information systems. However, understandings of data curation are necessarily emergent at this stage. Our case, through an ethnographic study of the actual practices of a pioneering data gathering and sharing effort in ecology, with more than 20 years of experience in data stewardship, is providing insights and understandings of the complexities and balances involved in data stewardship in one particular long-term program. We hope to raise some fundamental questions about the relationship between the actual work practices of data stewardship and the e-Science vision of data curating and sharing as well as to enrich the prevalent understanding of data curation.

For instance, with regard to increasingly automated data collection, there is the question of whether it represents a transition from or rather an augmentation of ongoing manual techniques. This cannot necessarily be known in advance and may require complex negotiations and alignments between related areas over extended periods of time to find the suitable combinations, configurations and emphases. While intimate relationships and intertwined change processes between science, data and technology are an established practice in LTER, the relationships between data curation, grid enabled technologies and e-Science environments seem unarticulated with development in each proceeding in a somewhat detached manner. To give an example, while there is a budding awareness of long-term concerns related to data curation in e-Science, the issue of technological long-term sustainability has been given limited consideration. Furthermore, whereas the LTER case study emphasizes the embedded position of data stewardship in the local science environments and illustrates how LTER information managers as practitioners of data managing and information infrastructuring have provided a solid and dependable foundation for addressing, evaluating and promoting technological visions in their particular contexts of science conduct, e-Science literature relates data curation to generic curation models and to large centralized facilities, such as data curation centers, that are removed from the actual conduct of science. Though the LTER community approach has not produced immediate or generic "solutions" to data management and federation, the networking of sites has generated broad ranging dialogue, much needed problem definition, valuable working solutions and technology arrangements, and new engagements with scientists. These arrangement have created a general "information management preparedness" and a "data stewardship awareness".

We suggest that the view of data curation in e-Science could be enriched by more accurate understandings of and closer connections with ongoing, situated data curation and stewardship practices in various settings of scientific collaboration. Attention directed to studying the concrete ways of conducting science, curating data and the complicated relations of data in their environments of scientific (re)use and curation/management holds the potential of providing more consistent understandings of existing and emerging data curation and stewardship practices. e-Science data curation would also benefit from involving science communities at the ground level from the beginning, for instance, by establishing mutually fruitful collaborative partnerships with such communities. In recognition of the view that partnerships are highly valuable in developing e-Science technologies and solutions, the ideal position and experience of embedded information managers can be used to mediate between domain scientists and e-Science technology developers.

Curation and stewardship both focus on the data but have different views about the nature of data, their life cycles and relations with their environments of science conduct. As portrayed in e-Science literature data curation in organizing and overseeing data holdings deals with guidelines and procedures for data ingestion, archive, and delivery. Data stewardship as practiced in LTER provides a large conceptual framework, an overarching process occurring now but attending to the past and taking into account and influencing the future, stretching from data planning to sampling, from data archive to use and reuse – including both data care and information infrastructure work. Such work involves data definitions, data requirements, and quality assurance as well as user feedback, redesign, and data exchange.

A decade of LTER conceptual work and field observations culminated in an articulation of the concept of ''the invisible present'' (Magnuson, 1990) capturing the notion of the past being present through slow processes and lagged results that have shaped today's ecosystems: ''The invisible present is the time scale within which our responsibilities for planet earth are most evident. Within this time scale, ecosystems change during our lifetimes...'' (Magnuson, 1990, p. 495). The phrase ''the invisible present'' also succeeds in conveying how time-series data bring the past forward to the present, a kind of telescoping of time. In a similar manner, the present projects into the future and thus engenders responsibilities and challenges for data stewardship. How do we handle the intertwined nature of past, present, and future as recorded by data records? The multiple dimensions of long-term time require sensitivity to care, continuity and competence while demanding a fuller understanding of the multiple dimensions of infrastructure.

Creating an acknowledgement of and a sensitivity to the long-term requires procedures, mechanisms and strategies be continually articulated and responsively designed and deployed. In terms of ''care'', methods are required to raise awareness of the recurring work of balancing, aligning and negotiating. Data care manifests itself in tending to community resources while ''continuity'' provides resilience for handling long-term data sets in dynamic research environments. It provides possibilities for developing an understanding of and attending to unavoidable tensions and conflicts whether in balancing multiple timeframes or local–global options. Where an informed ''competence'' is needed, roles are emerging: professionals whether designated information managers, data curators, or data scientists are working along side domain scientists/researchers and technicians. The work of tying together represents an expanded notion of information infrastructure that includes community processes that support the flow of data through local, community, and global arenas – the processes of data care, continuity, and competence over the long-term.

While e-Science and CSCW are synergistic in that they share similar research interests in technologically mediated scientific collaborations and they have complementary areas of expertise, they both lack an understanding of the long-term perspective and the multiplicity of temporal scales. With the pervasive call to meet short-term science needs, there is an ever-present danger that the long-term perspective will remain marginalized. Therefore we encourage research efforts on all fronts that explore further what is at stake with the long-term perspective.

Acknowledgements

This work is partially supported by an NSF/SBE/SES Human Social Dynamics grant #04-33369. The work is conducted in collaboration with the LTER community (NSF/OCE #04-17616, NSF/OPP #02-17282 and #04-05069). The fieldwork was conducted in 2002, and we offer our special thanks to Geoffrey C. Bowker for collaboration in the BDEI project (NSF/DGO #EIA-01-31958). Furthermore, we thank the anonymous reviewers for their constructive comments.

References

Ackerman, M. S. and C. Halverson (2004): Organizational Memory as Objects, Processes, and Trajectories: An Examination of Organizational Memory in Use. *Computer Supported Cooperative Work (CSCW), The Journal of Collaborative Computing*, vol. 13 no. 2, pp. 155–189.

Arzberger, P., P. Scroeder, A. Beaulieu, G. Bowker, K. Casey, L. Laaksonen, D. Moorman, P. Uhlir and P. Wouters (2003): *Promoting Access to Public Research Data for Scientific, Economic and Social Development*, OECD Follow Up Group on Issues of Access to Publicly Funded Research Data, Final Report. Available: http://www.dataaccess.ucsd.edu/Final_Report_2003.pdf [Last referenced: 23.05.2006].

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 Cyberinfrastructure*, Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure [Web-document]. Available: http://www.communitytechnology.org/nsf_ci_report/ [Last referenced: 23.05.2006].

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*, vol. 50 no. 11, pp. 963–978.

Baker, K.S., D. Ribes, F. Millerand and G.C. Bowker (2005): Interoperability Strategies for Scientific Cyberinfrastruture: Research and Practice. American Society for Information Systems and Technology. In *05ASIST. American Society of Information Science and*

Technology, Proceedings Bringing Research and Practice Together, Charlotte, North Carolina, October 28 to November 02, 2005.

Bertelsen, O.W. and S. Bødker (2001): Cooperation in Massively Distributed Information Spaces. In W. Prinz, M. Jarke, Y. Rogers, K. Schmidt and V. Wulf (eds.): *ECSCW. Seventh European Conference on Computer-Supported Cooperative Work, September 16 to 20, 2001* Dordrecht, Bonn, Germany: Kluwer Academic Publishers, pp. 1–17.

Birnholtz, J.P. and M.J. Bietz (2003): Data at Work: Supporting Sharing in Science and Engineering. In M. Tremaine (ed.): *GROUP'03. Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work, 2003 November 9 to 12, 2003*. ACM Press, pp. 339–348.

Bowker, G.C. (2000): Biodiversity Datadiversity. *Social Studies of Science*, vol. 30 no. 5, pp. 643–683.

Brand, S. (1994): *How Buildings Learn. What Happens After They're Built*. New York: Viking, pp. 243.

Buneman, P., L. Lyon and C. Rusbridge (2005): *Comments from the Digital Curation Centre on Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century, a draft report of the National Science Board*. Available: http://www.dcc.ac.uk/docs/nsbreport.pdf [Last referenced: 23.05.2006].

Callahan, J.T. (1984): Long-Term Ecological Research. *BioScience*, vol. 34 no. 6, pp. 363–367.

CCSDS 650.0-B-1 (2002): *Reference Model for an Open Archival Information System (OAIS)*. Washington, DC, USA: National Aeronautics and Space Administration.

Chervenak, A., I. Foster, C. Kesselman, C. Salisbury and S. Tuecke (1999): The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets. *Journal of Network and Computer Applications*, vol. 23 no. 3, pp. 187–200.

Chin, G. Jr. and C.S. Lansing (2004): Capturing and Supporting Contexts for Scientific Data Sharing via the Biological Sciences Collaboratory. In *CSCW'04. ACM Conference on Computer Supported Cooperative Work, November 6 to 10, 2004*. Chicago, Illinois, USA, pp. 409–418.

Christiansen, E. (1997): Gardening: A Metaphor for Sustainability in Information Technology-Technical Support. In J. Berleur and D. Whitehouse (eds.): *An Ethical Global Information Society: Culture and Democracy Revisited*, London: Chapman & Hall.

Dittrich, Y., S. Eriksén and C. Hansson (2002): PD in the Wild: Evolving Practices of Design in Use. In T. Binder, J. Gregory and I. Wagner (eds.): *PDC'02. Proceedings of the Participatory Design Conference, Malmö, Sweden, June 23 to 25, 2002*. Palo Alto, CA: CPSR, pp. 124–134.

Driscoll, C.T., G.B. Lawrence, A.J. Bulger, T.J. Butler, C.S. Cronan, C. Eagar, K.F. Lambert, G.E. Likens, J.L. Stoddard and K.C. Weathers (2001): Acidic Deposition in The Northeastern United States: Sources and Inputs, Ecosystem Effects, and Management Strategies. *BioScience*, vol. 51, pp. 180–198.

Finholt, T.A. and G.M. Olson (1997): From Laboratories to Collaboratories: A New Organizational Form for Scientific Collaboration. *Psychological Science*, vol. 8 no. 1, pp. 28–36.

Fisher, G. and J. Ostwald (2002): Seeding, Evolutionary Growth, and Reseeding: Enriching Participatory Design with Informed Participation. In T. Binder, J. Gregory and I. Wagner (eds.): *PDC'02. Participatory Design Conference, Malmö, Sweden, June 23–25, 2002* Palo Alto, CA: CPSR, pp. 135–143.

Fischer, G., E. Giaccardi, Y. Ye, A.G. Sutcliffe and N. Mehandjiev (2004): Meta-design: A manifesto for End-User Development. *Communications of the ACM*, vol. 47 no. 9, pp. 33–37.

Franklin, M., A. Halevy and D. Maier (2005): From Databases to Dataspaces: A New Abstraction for Information Management. *SIGMOD Record*, vol. 34 no. 4, pp. 27–33.

Gray, J., D.T. Liu, M. Nieto-Santisteban, A. Szalay, D.J. DeWitt and G. Heber (2005): Scientific Data Management in the Coming Decade. *SIGMOD Record*, vol. 34 no. 4, pp. 34–41.

Greenbaum, J.M. and M. Kyng (1991): *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Greif, I. and S. Sarin (1987): Data Sharing in Group Work. *ACM Transactions on Office Information Systems*, vol. 5 no. 2, pp. 187–211.

Grimm, N.B. and C.L. Redman (2004): Approaches to The Study of Urban Ecosystems: The Case of Central Arizona-Phoenix. *Urban Ecosystems*, vol. 7, pp. 199–213.

Gross, K.L., C.E. Pake, E. Allen, C. Bledsoe, R. Colwell, P. Dayton, M. Dethier, J. Helly, R. Holt, N. Morin, W. Michener, S.T. A. Pickett and S. Stafford (1995): *Final Report of the Ecological Society of America Committee on the Future of Long-term Ecological Data (FLED)*, Volume I: Text of the Report. Washington, DC: The Ecological Society of America.

Hanseth, O., E. Monteiro and M. Hatling (1996): Developing Information Infrastructure: The Tension between Standardisation and Flexibility. *Science, Technology & Human Values*, vol. 21 no. 4, pp. 407–426.

Harmon, M.E., K.J. Nadelhoffer and J.M. Blair (1999): Measuring Decomposition, Nutrient Turnover, and Stores in Plant Litter. In G.P. Robertson, C.S. Bledsoe, D.C. Coleman and P. Sollins (eds.): *Standard Soil Methods for Long Term Ecological Research* New York: Oxford University Press, pp. 202–240.

Hayden, B.P. (2000): Climate Change and Exratropical Storminess in the United States: An Assessment. *Journal of American Water Resources Association*, vol. 35 no. 6, pp. 1387–1397.

Hedstrom, M. (2003): *It's About Time: Research Challenges in Digital Archiving and Long-term Preservation, Final Report*. Workshop on Research Challenges in Digital Archiving and Long-term Preservation, April 12 to 13, 2002. Sponsored by the National Science Foundation and The Library of Congress.

Helly, J.J., T. Todd Elvins, D. Sutton, D. Martinez, S.E. Miller, S. Pickett and A.M. Ellison (2002): Controlled Publication of Digital Scientific Data. *Communications of the ACM*, vol. 45 no. 5, pp. 97–101.

Henderson, A. and M. Kyng (1991): There's No Place like Home: Continuing Design in Use. In J. Greenbaum and M. Kyng (eds.): *Design at Work*, London, New Jersey: Lawrence Erlbaum.

Hey, T. and A.E. Trefethen (2003): The Data Deluge: An e-Science Perspective. In F. Berman, G. Fox and T. Hey (eds.): *Wiley Grid Computing: Making the Global Infrastructure a reality*. John Wiley & Sons Ltd., pp. 809–824.

Hilgartner, S. (1995): Biomolecular Databases: New Communication Regimes for Biology?. *Science Communication*, vol. 17 no. 2, pp. 240–263.

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*, vol. 53 no. 1, pp. 21–32.

Hodge, G. and E. Frangakis (2004): *Digital Preservation and Permanent Access to Scientific Information: The State of the Practice (CENDI/04–3)*, The International Council for Scientific and Technical Information (ICSTI) and CENDI (U.S. Federal Information Managers Group). February 2004, Revised April 2004. Available: http://www.icsti.org/icsti/icsti_reports.html [Last referenced: 23.05.2006].

Jirotka, M., R. Procter, M. Hartswood, R. Slack, A. Simpson, C. Coopmans, C. Hinds and A. Voss (2005): Collaboration and Trust in Healthcare Innovation: The eDiaMoND Case

Study. *Computer Supported Cooperative Work (CSCW), The Journal of Collaborative Computing*, vol. 14 no. 4, pp. 369–398.

Johansen, R. (1988): *Groupware. Computer Support for Business Teams*. New York: The Free Press.

Kanstrup, A.M. (2005): *Local Design: An Inquiry into Work Practices of Local IT Supporters*. PhD theses. Department of Communication, Aalborg University, Denmark.

Kaplan, S. and L. Seeback (2001): Harnessing Complexity. In *ECSCW. Proceedings of the Seventh European Conference on Computer Supported Cooperative Work, September 16 to 20, 2001, Bonn, Germany*. Netherlands: Kluwer Academic Publishers, pp. 359–397.

Karasti, H. and K.S. Baker (2004): Infrastructuring for the Long-Term: Ecological Information Management. In *HICSS'3. Proceedings of the Hawaii International Conference on System Sciences 2004*, Hawaii, USA, January 5 to 8, 2004.

Karasti, H. and A.-L. Syrjänen (2004): Artful Infrastructuring in Two Cases of Community PD. In *PDC 04. Proceedings of the Eighth Conference on Participatory Design: Artful integration: interweaving Media, Materials and Practices, Volume 1, Toronto, Ontario, Canada, July 27 to 31, 2004*. New York: ACM Press, pp. 20–30.

Lamb, R. and E. Davidson (2005): Information and Communication Technology Challenges to Scientific Professional Identity. *The Information Society*, vol. 21 no. 1, pp. 1–24.

Lave, J. and E. Wenger (1991): *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

Likens, G.E. (1989): *Long-Term Studies in Ecology: Approaches and Alternatives*. New York: Springer-Verlag.

Lord, P. and A. Macdonald (2003): *e-Science Curation Report-Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision*. Prepared for the JISC Committee for the Support of Research (JCSR). Twickenham, UK, The Digital Archiving Consultancy Limited. Available: http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf [Last referenced: 23.05.2006].

Lord, P., A. Macdonald, L. Lyon and D. Giarretta (2004): From Data Deluge to Data Curation. In *Proceedings of the UK e-science All Hands meeting 2004*, pp. 371–375.

Magnuson, J.J. (1990): Long-Term Ecological Research and the Invisible Present. *BioScience*, vol. 40 no. 7, pp. 495–501.

Magnuson, J.J., D.M. Rogbertson, B.J. Benson, R.H. Wynne, D.M. Livingsone, T. Arai, R.A. Assel, R.G. Barry, V. Card, E. Kuusisto, N.G. Granin, T.D. Prowse, K.M. Stewart and V.S. Vuglinski (2000): Historical Trends in Lake and River Ice Cover in The Northern Hemisphere. *Science*, vol. 289, pp. 1743–1746.

Markus, L.M. (2001): Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success. *Journal of Management Information Systems*, vol. 18 no. 1, pp. 57–93.

Michener, W.K. (2006): Meta-Information Concepts for Ecological Data Management. *Ecological Informatics*, vol. 1, pp. 3–7.

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*, vol. 7 no. 1, pp. 330–342.

National Science Board (2005): *Long Lived Digital Data Collections: Enabling Research and Education in the 21st Century*, National Science Board (NSB-05-40, Revised May 23, 2005). Available: http://www.nsf.gov/pubs/2005/nsb0540/ [Last referenced: 23.05.2006].

Newman, H.B., M.H. Ellisman and J.A. Orcutt (2003): Data-Intensive e-Science Frontier Research. *Communications of the ACM*, vol. 46 no. 11, pp. 68–77.

OECD Global Science Forum (2005): *Organisation for Economic Co-operation and Development Global Science Forum Report on Grids and Basic Research Programmes*. Final consensus report from the OECD Global Science Forum Workshop, Sydney, Australia, September 25–27, 2005.

O'Day, V.L., A. Adler, A. Kuchinsky and A. Bouch (2001): When Worlds Collide: Molecular Biology as Interdisciplinary Collaboration. In W. Prinz, M. Jarke, Y. Rogers, K. Schmidt and V. Wulf (eds.): *ECSCW. Seventh European Conference on Computer-Supported Cooperative Work, September 16 to 20, 2001, Bonn, Germany*. Netherlands: Kluwer Academic Publishers, pp. 399–418.

Pickett, S.T.A., W.R. Burch and J.M. Grove (1999): Interdisciplinary Research: Maintaining the Constructive Impulse in a Culture of Criticism. *Ecosystems*, vol. 2, pp. 302–307.

Pipek, V. (2005): *From Tailoring to Appropriation Support: Negotiating Groupware Usage. Doctoral thesis*. Acta Universitatis Ouluensis, Series A, Scientiae rerum naturalium nro 430. Oulu 2005.

Rolland, K.H. and E. Monteiro (2002): Balancing the Local and the Global in Infrastructural Information Systems. *The Information Society*, vol. 18, pp. 87–100.

Rolland, K.H., V. Hepsø and E. Monteiro (2006): (Re)Conceptualizing Common Information Spaces across Heterogeneous Contexts: Im/Mutable Mobiles and Imperfection. Accepted for *CSCW'06. ACM Conference on Computer Supported Cooperative Work*.

Sandusky, R.J. (2003): Infrastructure Management as Cooperative Work: Implications for Systems Design. *Computer Supported Cooperative Work (CSCW), The Journal of Collaborative Computing*, vol. 12, pp. 97–122.

Schmidt, K. and L. Bannon (1992): Taking CSCW Seriously: Supporting Articulation Work. *Computer Supported Cooperative Work (CSCW), The Journal of Collaborative Computing*, vol. 1 no. 1–2, pp. 7–40.

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

Sonnenwald, D.H. (2003): Expectations for a Scientific Collaboratory: A Case Study. In *GROUP '03. Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work 2003. November 9 to 12, 2003*. Sanibel Island, Florida, USA, pp.68–74.

Star, S.L. and G.C. Bowker (2002): How to Infrastructure. In L.A. Lievrouw and S. Livingstone (eds.): *Handbook of New Media: Social Shaping and Consequences of ICTs* London: SAGE Publications, pp. 151–162.

Star, S.L. and J.R. Griesemer (1989): Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, vol. 19, pp. 387–420.

Star, S.L. and K. Ruhleder (1996): Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, vol. 7, pp. 111–133.

Star, S.L. and A. Strauss (1999): Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW), The Journal of Collaborative Computing*, vol. 8 no. 1–2, pp. 9–30.

Sterling, T.D. and J.J. Weinkam (1990): Sharing Scientific Data. *Communications of the ACM*, ACM Press, vol. 33, no. 8, pp. 112–119.

Strauss, A.L. (1975): *Chronic Illness and the Quality of Life*. Saint Louis: The C. V. Mosby Company.

Strauss, A.L., S. Fagerhaugh, B. Suczek and C. Wiener (1985): *Social Organization of Medical Work*. Chicago: University of Chicago Press.

Suchman, L. (1995): Special Issue: Representations of Work. *Communications of the ACM*, vol. 38 no. 9, pp. 33–68.

Suchman, L. (2000): Located Accountabilities in Technology Production. Work-in-progress, revision of (Suchman, 1994), presented at the *Sawyer Seminar on Heterarchies, Santa Fe Institute, October 2000*.

Suchman, L., J. Blomberg, J.E. Orr and R. Trigg (1999): Reconstructing Technologies as Social Practice. *American Behavioral Scientist*, vol. 43 no. 3, pp. 392–408.

UK Research Council e-Science definition (2001): Available: http://www.rcuk.ac.uk/escience/. [Last referenced: 23.05.2006].

Van House, N.A., M.H. Butler and L.R. Schiff (1998): Cooperative Knowledge Work and Practices of Trust: Sharing Environmental Planning Data Sets. In *CSCW '98. Proceedings of the ACM Conference On Computer Supported Cooperative Work, November 14 to 18, 1998*. Seattle, WA: ACM, pp. 335–343.

Waide, R.B., M.R. Willig, C.F. Steiner, G. Mittelbach, L. Gough, S.I. Dodson, G.P. Juday and R. Parmenter (1999): The Relationship between Productivity and Species Richness. *Annual Review of Ecology and Systematics*, vol. 30, pp. 257–300.

Zimmerman, A.S. (2003): *Data Sharing and Secondary Use of Scientific Data: Experiences of ecologists*. Ph.D. Dissertation, University of Michigan.