J Intell Inf Syst (2007) 29:127–144
DOI 10.1007/s10844-006-0035-7

Information ecology: open system environment for data, memories, and knowing

Karen S. Baker · Geoffrey C. Bowker

Published online: 28 February 2007

Abstract An information ecology provides a conceptual framework to consider data, the creation of knowledge, and the flow of information within a multidimensional context. This paper, reporting on a 1 year project to study the heterogeneity of information and its management within the Long Term Ecological Research (LTER) community, presents some manifestations of traditionally unreported 'invisible work' and associated elements of informal knowledge and unarticulated information. We draw from a range of ethnographic materials to understand ways in which data-information-knowledge are viewed within the community and consider some of the non-linear aspects of data-knowledge-information that relate to the development of a sustained, robust, persistent infrastructure for data collection in environmental science research. Taking data as the unit of study, the notion of long-term research and data holdings leads to consideration of types of memory and of knowledge important for design of cyberinfrastructures. Complexity, ambiguity, and nonlinearity are part of an information ecology and addressed today by exploring multiple types of knowledge, developing information system vocabularies, and recognizing the need for intermediation.

Keywords Memory · Infrastructure · Information ecology · Data management · Long-term

1 Introduction

This paper develops insights gained from a project that brought together an interdisciplinary team to conduct a 1 year joint study of information management within the Long Term Ecological Research (LTER) community. We discuss memory and its relationships to data, information, and knowledge. Memory practices (Bowker, 2006) are at the center of LTER work—the community is aiming to build very long baselines of environmental data, base-

K. S. Baker (✉)
University of California, San Diego, CA, USA
e-mail: kbaker@ucsd.edu

G. C. Bowker
Santa Clara University, Santa Clara, CA, USA
e-mail: gbowker@scu.edu

lines suited to the life of the ecosystem rather than (as is currently the case) to the lifetime of the researcher. We analyze our ethnographic materials to draw out ways that knowledge is held in the LTER community. Such understanding informs information systems' design and impacts development of a robust, persistent infrastructure supporting data work and scientific practices in environmental science. We present a conceptual framework for an information ecology inclusive of data sets and data collectors, information systems and knowledge makers, as well as digital federations and social networks. The framework is associated organizationally with local data centers, community learning centers, and global grids, respectively.

This work was carried out as part of the NSF Biodiversity and EcoInformatics project entitled 'Designing an Infrastructure of Heterogeneity in Ecosystem Data, Collaborators and Organizations.' Our interdisciplinary team working collaboratively at the interface of environmental sciences, social sciences, and information sciences (Baker, Bower, & Karasti, 2002) was comprised of an LTER information manager, a science and technology studies expert from the field of communication, an ethnographically trained information systems designer, and the LTER community. Ethnographic fieldwork consisted of participant observations, transcribed interviews, and focused visits to sites, meetings, and workshops. Paper and digital documents as well as photos were collected across all major roles and categories (site, network, and information management).

2 The ecology of long term databases

The database is the cultural and technoscientific object of our times: as rich in its implications as has been the cinema and the printed book (Manovich, 1999). Babbage (1837) wrote that the invention of printing had taken us from being blind creatures of instinct precisely through its constitution of a prosthetic memory. However, that prosthetic memory was very limited. There was little ability to randomly access books or parts of books without an enormous labor of very imperfect cataloging, indexing and abstracting. There were few physical copies of information held in books, so lumbering mountains would have to journey across Europe to meet their leather bound Mahomets. With digital databases, we are reconstituting our science, government and arts.

And yet the data in databases never stands alone. As Walsh and Ungson (1991) pointed out in their classic text on organizational memory, there are several different 'containers' for memory in an organization—and these interoperate. Not all information has to be recorded in digital or other archival form. Consider the total institution. Douglas (1986) argues that: "when everything is institutionalized, no history or other storage devices are necessary." If I get processed into a prison, I can survive there as just a number (as the Count of Monte Cristo discovered). There is no need for the institution to hold any information about me other than that I exist and that I am subject to its regulations for such a time period; there is no need for me to remember anything about my own past, or any sets of skills beyond a fairly simple motor set. Why I am there and who I am just don't matter to the institution itself—it 'remembers' all it needs to know through the complex set of procedures that it puts into place. Contrast this extreme example with participant roles in a research community: how much is documented in a job description versus developed in practice; how much is recorded as accomplishments versus accumulated in experiences which in the case of the LTER community incorporates multi-task roles, cross-site research activities, and interdisciplinary meetings in addition to cooperative environmental field studies. Articulating context introduces one aspect of complexity (Kaplan & Seebeck, 2001; Weick & Sutcliffe, 2001) suggesting that a research community member or a database entry may be wrapped in only a partially articulated context. Multiple perspectives and criteria introduce a complexity that unfolds into an ambiguity of solutions (Smith & Marx, 1994; Star & Ruhleder, 1996) and nonlinearity of developments (Solomon, 1997; Spasser, 1997; Wauzzinski, 2001; Yates-Mercer & Bawden, 2002).

The replacement of memory by procedures extends to a formal information processing argument that Ashby (1956) made about closed systems of all kinds. He argued that if we completely know a system in the present, and we know its rules of change (how a given input leads to a given output) then we don't need to bring to mind anything about the past. Memory, he said, is a metaphor needed by a 'handicapped' observer who cannot see a complete system, and "the appeal to memory is a substitute for his inability to observe...." Now no institution is ever total, nor is any system totally closed. However, it remains true that there are modes of remembering that have very little to do with consciousness on the one hand or formal recording keeping on the other.

Traces are physical evidences of actions and thoughts. The ecology of memory traces is something we live on a daily basis. It is rarely theorized as an ecology—more often it is given completely differential value: memory held in the head is just not the same sort of thing as memory held in a file cabinet (Hutchins, 1995). Such a differentiation has some heuristic value for us in that it permits several discipline-bound investigations of memory; but it has little grounding—it is like the separate 'pots' of money ('rainy day'; 'college fund' and so forth) that people create to differentiate their undifferentiable supply of money (it's just a number, after all). Its limited heuristic value is accompanied by its negative consequence of forcing us into a fractured view of our memory work.

3 Long term ecological databases

For much contemporary scientific work, data reuse has become a clarion call while also raising questions (Zimmerman, 2003). We are currently across the board creating petabits of data—be they streaming satellite images of our world, probes into space, remote sensors embedded in the wilderness or seismic data echoing from the deep mantle. There is a lot more data being produced than there are scientists and techniques to process them. Further, there is a convergence operating between many sciences around issues of global concern such as the thinning of the ozone layer, climate change, and habitat preservation. The list has been ever lengthening since the turn of the nineteenth century (Ecological Visions Committee, 2004; NRC, 2001; Serres, 1990). In order to answer the questions that the world is posing, scientists require interoperable databases—which package time, space, quantity and type in mutually comprehensible ways. They need to be able to share, or at least negotiate protocols between, multiple ontologies.

In ecological science, the challenges of unifying time scales, agreeing on spatial units, and clarifying species lists are staggering. Central to long-term ecological studies is the ambition of producing ecological data for the ages (Likens, 1989; Magnuson, 1990). Ecosystems do not develop in 30 year chunks (the average career length of an environmental scientist). And yet environmental data has in general been collected by a few scientists at most collaborating together for the fixed period of a grant or project. When data collectors retire, typically their protocols are not sufficiently well enumerated for future generations to use the data (Bowser, 1986). Even when they are well enough preserved, there is the tendency to gather new data with new tools rather than rework old data. There are at least two prompts for this: one is a feature of professional structures and personal inclinations, which push scientists toward the latest technology; another is a fuzzier notion of a learning process, which occurs while scientists are making observations and measurements.

Facing the scaled-up global task of databasing knowledge about the environment, we confront directly the past of environmentally related sciences. There are huge species lists drawn up at Kew Gardens in England and at Harvard in the United States (not to mention lists from non-Anglophone countries). In the nineteenth century, when these lists were first developed, there was relatively little need to reconcile the respective nomenclatures. Now as our scientific and social concerns are becoming global in scope, and as plants themselves are traveling the globe, it is increasingly making a difference whether or not a given designation in Ireland is really or not the same as one in New Zealand. In the past, such questions have been posed on an ad hoc basis, with taxonomists traveling from botanical garden to botanical garden or receiving type specimens of plants through the mail. It has been estimated that the rate of synonymy (the same plant having different names) is of the order of 20% across these lists. Even with these two lists reconciled, there remains the work of joining together all of the local lists that draw from different editions of different standard works in the field (for example, GAP analysis in the United States is held to very different State species lists; since State policy for protection of species responds to the local list (Edwards et al., 1995). Even if all the names were agreed upon, it is extremely difficult to conjure information into the right spatiotemporal units. The older data becomes, the more variable the units; however even recent data comes in a staggering variety of forms. Today, technology is enabling the sharing of lists and their organizational structure through internet presentations such as the International Catalogue of Life Programme dynamic list checklist (COL; http://www.sp2000.org/dynamicchecklist.html) and the Global Biodiversity Information Facility search (GBIF; http://www.gbif.org/portal).

The Long-Term Ecological Research (LTER) program consists of 26 research teams of investigators, each team investigating a particular site's biome in a defined study region (Franklin, Bledsoe, & Callahan, 1990; Hobbie, 2003). Each team works independently to understand the ecology of their locale; each team also works collectively on cross-site themes and activities. This federation of independent research sites works with a network office and a sense of community. There is a continuity of program funding in renewable 6-year cycles, which creates a stable environment promoting cooperation and enabling innovative test bed activities. Data management has been a required part of each site's research program since LTER began in 1980 supported by National Science Foundation.

Goals of the LTER are to carry out a series of very long-term measurements using documented protocols, to incorporate data management as part of the scientific work itself, and to promote dialogue through shared activities. Here, then, there is a focus on creating databases, which will be useful for very many years, and ultimately across multiple scientific boundaries. Which brings us directly to the question of metadata. A standard response to the difficulty of creating interoperable databases today is to agree on a set of metadata standards. In a sense, however, the whole problem just recurses here—since there is a proliferation of metadata standards within environmental science as significant as the proliferation of data standards themselves. There is little historical evidence that the branching data and metadata standards can be stopped. This suggests the need to accept that there are very real social, organizational and cognitive machineries of difference, which continually fracture standards into local versions (Boland & Tenkasi, 1995; Bowker, 2006). Rather than see this as a problem to be overcome solely by another new and better standard, our study of information management within the LTER leads us to propose that a careful analysis of the political and organizational economy of memory practices in interdisciplinary environmental science may lead to the development of new perspectives in very long term information management.

4 LTER case study: an integrative perspective

The field of ecology stresses the links and associations within a system as much as the differences and dominions so presents a multifaceted approach to interdependencies of environmental, human, and technological factors, including explorations of principles of self regulation and self correction. The LTER provides an interdisciplinary laboratory with participants accumulating shared experiences (Greenland, Goodin, & Smith, 2003; Kinzig et al., 2000; Robertson, Coleman, Bledsoe, & Sollins, 1999). It provides a sheltered forum in which to explore information management grounded within a scientific program and to consider the meanings and impacts of interdisciplinarity, data sharing, and technology use on the work of long-term research (Baker et al., 2000; Karasti & Baker, 2004; Michener, Brunt, & Stafford, 1994; Pickett & Cadenasso, 2002; Redman, Grove, & Kuby, 2000). It provides, in addition, an arena in which to consider the ramifications of digitally constructed memories for domain and cross-domain knowledge bases as data and documents are gathered into global collections with negotiated structures and standardized classifications.

With the LTER Information Management Committee recent adoption of the Ecological Metadata Language (EML, Michener, Brunt, Helly, Kirchner, & Stafford, 1997) as a metadata standard, a new community endeavor focused on implementation of the standard has been initiated. The integration and bridge of local conventions and work practices to community standards is an ongoing activity creating valuable opportunities for articulation of the often underestimated ramifications of such a process.

4.1 Multiple dimensions of an information ecology

Traditionally a simplified view of a scientist's role might be described as twofold: to do good science and to record knowledge gained in peer reviewed publications. With the concept of a long-term science program or of long-term data, new expectations and work practices arise. Amidst contemporary calls to "give me the data" and "show me the knowledge," research community participants endeavor to address the data-information-knowledge trinity, its elements and relationships. This trinity suggests a linear progression with knowledge appearing as an end of the pipeline product, the end result of a unidirectional filtering process: data that is filtered into information that is filtered into knowledge (Poore, 2003). Alternative perspectives on differences and relationships can be explored using a four-paradigm grid (Deetz, 1996). Such a two-dimensional quadrant approach has been used to explore dimensions of knowledge, such as knowledge holder and knowledge type, with axes of individual-group and explicit-implicit knowledge, respectively (Gasson, 2004; Jordan, 1996).

A data-knowledge grid is presented in Fig. 1; the horizontal dimension refers on the left to emergent knowledge at the local or ground level that is informed by working practices and experiences. To the right is a more explicit articulated form of knowledge. The vertical dimension differentiates data at an individual or local collection point and its flow into group, managed collections with documented structures and standardized classifications. Circular arrows indicate some flows of information: the transformation of data into information (designed to standards) that flows into a domain knowledge base

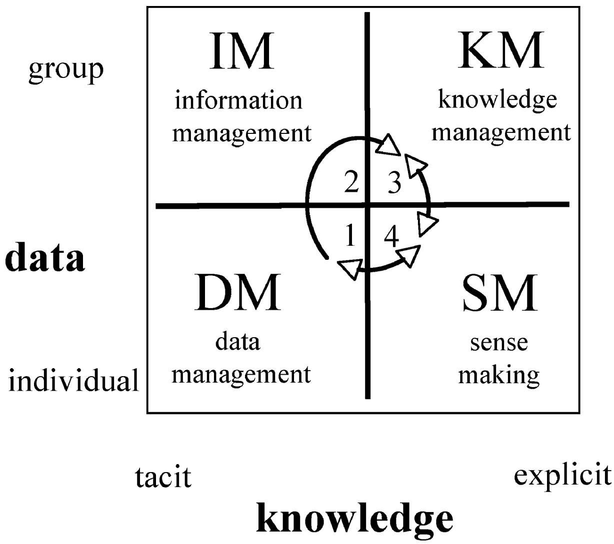

Fig. 1 A data-knowledge grid presents a range of arenas within which data is handled from individual to group together with a range of knowledge types from tacit to explicit

(designed for interoperability) is illustrated in the movement from quadrants 1 through 3. In quadrant 3, the domain level knowledge is acknowledged within the community as a recognized synthesis. In addition, local knowledge highlighting organizing (Weick et al., 2005) in quadrant 4 (sense making) is shown to influence the structure of domain knowledge (quadrant 3) as well as contributes to the discussion about data collection (quadrant 1).

There are dynamics of change operating within each quadrant and feedbacks across all the quadrant boundaries. Through the exchange and flow of information, an organization blends tacit and explicit knowledge to create new knowledge and plans for action. The result of activity in each of these quadrants is relevant, overlapping and yet distinct: The result of work in quadrant 1, with data at or close to its collection point, is the data itself. An LTER scientist recognizes the insight and ramifications of data handling as huge:

> Well, we care a lot about data management. From the outside, I think it has been a big focus ... and part of it is that it is not transparent, what you do as the data manager... that the way we handle information is going to have a great deal with what we end up saying that information has told us. And we have to be pretty insightful about that. And so from the beginning we saw data management as a huge aspect of the LTER and it has worked out far huger than probably one imagined.

and LTER information managers are explicit about the metadata of the data, a contemporary 'memory trace' of the data from the start of the data collection process:

> Because the data it, that is where it all starts, and the knowledge about the data.

> We don't accept data without metadata ... because we don't want useless data. It doesn't help having a whole lot of data if you don't have any documentation describing those datasets.

> we are actively working towards getting that knowledge about their data from them before we, you know, loose that data, they pass away or become ill, that type of thing.

Quadrant 2 is defined by the requisite characteristic of interoperability that is required for automated handling of data. An LTER information manager explains that scaled, intersite research requires comparable data:

> Modern information management has become crucial for intersite ecological research. Any scientist who has gathered data from diverse data sources has dealt with issues of making data comparable, dealing with multiple data formats, structuring the aggregated data in a form that facilitates answering research questions, and providing access to the data and derived data and documents to colleagues.

and an LTER scientist refers to the non-trivial, worrisome task of making data intercomparable:

> information management gives me a headache and so I just want to be able to do what I do and then have that be easy. So you know, for instance, for data comparability between sites, I just really would actually prefer someone else worry about that: what does it mean that we don't have the same resolution and what does it mean that it doesn't cover the same time scales?

The work of making data intercomparable is, as this latter quote indicates, highly problematic. It is essentially an altruistic act in a highly competitive world for a scientist to take the time to make their data usable by someone else; particularly when they do not necessarily anticipate any return on their investment in terms of reusable data from elsewhere (frequently the case for ecological data, where it is third party users—planners, development agencies, policymakers, modelers working in other arenas like climate change —who are beneficiaries of interoperability).

Quadrant 3 with global data and knowledge represents another scaling and a transformation which an information manager describes as a synthesis of knowledge sharing:

> The vision has been identified that IM will go to the next level of knowledge sharing and synthesis of that knowledge...while IM will continue to serve a custodial role for data, to protect it for possibility of prosperity and longevity...

and a scientist's view of the synthesis includes the outcome, a research paper:

> And the sites would bring information to one of these meetings, not data, information, their synthesis of their own site data and then we try to work out some way of analyzing the knowledge that we had across these sites, and come up with a synthesis paper.

Quadrant 4 is more elusive and less frequently articulated but perhaps the research proposal can be considered a physical manifestation of the work in this arena. To emphasize the dynamic nature of quadrant 4 and to counteract the concept of knowledge as a static end product by highlighting the process, the sense-making quadrant can also be labeled 'research knowing' (Boland & Tenkasi, 1995; Suchman, 2000; Whitley, 2000). Such knowing involves communications and often is grounded in shared field work:

> When it is most successful, this thing that has happened is some form of important communication that leads one person to say, gee that really helps me. What that discipline is relative to my discipline, there has to be some form of communication [for] that [to] happen. In our sciences it happens most frequently on field trips, or joint campaigns to go in the field to do some work, because you are away from telephones, you are away from everything, and you end up asking each other questions without any embarrassment whatsoever [about] not knowing something the other guy should know.

A scientist new to the LTER network refers to the size of the task in making interdisciplinarity part of local knowledge:

> I am really beginning to get more of a feel of the size of the intellectual challenge that is really involved with becoming more familiar with this large body of ecological theory and knowledge, and people who speak different vocabularies and have very different backgrounds and different ways of working and so I think that is going to be kind of fun.

Although artificial and incomplete, the quadrant diagram is a heuristic devised to prompt discussion. For instance, where do training and innovation occur? What relationships exist over the long-term between quadrants 3 and 4? When the focus is on data, discussion commences with data collection so quadrant 1 is labeled the first quadrant. If instead the research process were the element to be highlighted, the current quadrant 4 would be the quadrant considered first. Figure 1 provides a purposeful contrast with the traditional data-information-knowledge linear construct. Flows across quadrant boundaries are suggestive of iterative knowledge construction processes. In the rush to capture domain knowledge in digital form and to create information systems, there is a need to understand the complex flows of information that constitute an ecology so as to inform our choices of how to use technology and to enhance knowledge retention. It is important to model this process accurately in order to be able to determine suitable allocations of resources and of information roles.

4.2 Developing language to support the flow of information

The elements of language and vocabulary take on critical importance with the multiple types of data handling, information flow, and knowledge creation as well as the changing relationships between data, science, and technology. There are few cross-domain agreed upon standards for data, definitions of information, knowledge system frameworks, design strategies, or approaches to the tensions between the processes of science and the products of technology. When this is acknowledged as a dynamic 'state-of-our-art', an exciting research agenda opens up. On the other hand, with narrowly defined tasks, we risk putting in place static systems that create barriers to inquiry—Hughes (1983) demonstrates clearly how the development of large systems necessitates this boundary crossing. For example, consider a project defined by technical requirements alone and contrast this with a comprehensive analysis framework such as that summarized in Table 1 (Iivari, 1991). Here the concept of system design is broad, covering methodology and ontology as well as aspects of ethics and epistemology. Stewardship of work within an information ecology and for building a sustainable cyberinfrastructure requires ongoing discussions of a full suite of information system elements, their ramifications and interdependencies. By way of example, we present the categories from Iivari in Table 1 with the added element 'multi-perspectivism' in the epistemology category in order to broaden the table and to prompt consideration of additional alternatives (Chalmers, 1976).

The LTER network is organizationally federated through annual scientific and IM meetings, periodic All Scientists meetings, cross-community committees, shared research themes, and shared experiences from conference calls to LTER review panel participation. Although individual sites make choices with respect to using or not using particular technologies in support of local research, the Network Office information practices are cast more in the role of exploring and using technological tools as the means to support a digitally federated network. Federation is supported by infrastructure elements such as the LTER Network Information System (NIS) and the NIS Advisory Committee (NISAC). The concept of the Network

Table 1 An information system analysis framework

Methodology

Constructive methods
  *conceptual development
  *technical development
  *triangulation
Nomothetic methods
  *formal-mathematical
  *experiments
  *field studies/surveys
Idiographic methods
  *case studies
  *action research

Ethics

Role of IS science
  *means-end oriented
  *interpretive
  *critical
Values of IS Research
  *org/mgmt oriented
  *user oriented
  *others (educative)

Epistemology

  *positivism
  *anti-positivism
  *multiperspectivism

Ontology

View of information/data
  *descriptive facts
  *constitutive meanings
View of information/data system
  *technical system
  *organization/social system
View of human beings
  *determinism
  *voluntarism
View of technology
  *technological determinism
  *human choice
View of organization and society
  *realism
  *structuralism
  *interactionism
  *nominalism

After Iivari 1991; Karasti 1994

Information System (NIS) is undergoing change. Originally viewed as a technical product comprised of modules, it is beginning to be recognized as a scientific process.

> So building the NIS, and I think the exec's recommendation to making this committee (NISAC) sort of indicates that we are all recognizing that the building of the NIS is as much of a scientific process as it is a data management process.

NIS discussions were spurred by an NSF mandate:

> Big breakthrough came (in) 1994, the coordinating committee, with substantial help and encouragement from NSF, mandated that each site should have at least one dataset online.

This mandate for online data was followed in the next years by design work, implementation plans, and site prototype modules (Baker et al., 2000, 2002; Brunt, 1998; Henshaw et al., 1998). The development of NIS administrative modules occurred over a period of years but as the need for development of cross-site data modules grew, a NIS Advisory Committee (NISAC) explicitly composed of both information managers and scientists was formed in the Fall of 2002.

The LTER NIS development has incorporated some of the elements in Table 1. Conceptual themes of long-term, community, and interdisciplinarity as well as methodological concerns with data arose frequently in LTER interviews. In addition, the ethics of data use and of collaboration were often articulated while issues such as community scaling, technological automation, organizational trust, and community success were not developed. Ensuring trust of data through data quality control measures was discussed often but not trust with respect to technology, collaboration, or knowledge. Mention of the concept of perspectives and of technical progress recurred in interviews while reference to epistemologies occurred once.

> There are some fundamental issues about the nature of knowledge—I mean epistemological issues. How do you know what is good science? And there is some real down to earth prejudice between disciplines and good science.

With respect to ontological elements, an ongoing community project is to ensure that data is well described through metadata. An interesting area for future study in 'ontological mismatches' between differing subdisciplines in environmental science—for example when different methods yield varying measurements for the same attribute such as when 'productivity' is measured or when different standards are taken for describing sampling locations or size classification schemes.

4.3 Long-term data costs

Data structures frequently are designed in response to short-term needs and/or profit strategies. Notions of long-term for digital data and memory are being explored within informatics. The implications of long-term science are being explored via new programs such as the LTER program in ecology (Kaiser, 2001), time-series programs in oceanography (e.g. Ohman & Venrick, 2003; US JGOFS, 2001), and collaboratories in a variety of fields (Finholt, 2002). Brand (1994) discusses the concept of long-term maintenance in architecture and outlines the costs that are cumulative in supporting the inevitable maintenance and modifications to buildings—the cost of cleaning the Pompidou Center in Paris is already several times that of building it. The situation is analogous with technology where there are rapid rates of change with hardware and software (building materials), data content and organization (furniture), as well as participant views and interests (values and styles). Only recently have we accumulated a pool of experiences at both personal and organizational levels with respect to data use and re-use (house purchase and remodel) which may provide an impetus to broaden of our views regarding technology costs and data planning (Eriksen, 2001). The multiple time frames involved in the arena of data management appear as a juggling of short-term and long-term concerns and are evident in the everyday practices of information managers (Karasti & Baker, 2004). There are important ramifications to handling a long-term resource like data in a culture oriented to short-term market thinking. Further, information has special characteristics in terms of value, consumption, life-cycle and individuality (Eaton & Bawden, 1991). An integrative approach is unfolding today in explorations of information ecologies (Davenport, 1997; Star & Strauss, 1999). The LTER community builds from this understanding, highlighting long-term aspects of ecological studies (Gosz, 1999; Hobbie, 2003) while the diversity of sites, participants, and approaches creates a presence of both elemental and system views of natural systems (Odum, 1995, 1998; Orr, 2002).

Drawing from Brand (1994), Fig. 2 presents conceptually a cumulative view of data costs over time with data collection initially the major expense. As expressed by a scientist familiar with the LTER philosophy of data management:

> The information management, they have a very, very good point, that it is not that that is an unfunded mandate, I mean everybody realizes that they have to do it, but it is just

costing more and more and more and (a funding agency) has not been providing more and more money specifically for doing that.

Figure 2 illustrates conceptually some of these costs such as the periodic need for attention to maintaining data in a repository or archive as systems change and information must be migrated to new structures. In addition, there is a constant expense over time for maintaining access to the data. Science budgets have struggled over time to reflect realistic plans and resources for data collection but understandings of long-term costs for data maintenance, access, and integration remain elusive.

5 How are the memories chosen? Who speaks for the data?

5.1 Information ecology—structures

Data centers, learning centers, and global networks all are terms describing dynamic, complex work environments. Traditionally distinct, the movement of data from local collection to global network accessibility, suggests a larger scale, more integrated view. Figure 3 presents a simplified information ecology with three interrelated components, each emphasizing different processes and products in the life cycle of data. A project, platform, or data center with one or more data systems, represents the location where data is received from the laboratory or field and where structuring occurs. At a community center, data is analyzed, procedures established, and misalignments identified. This is a learning center, representing the point at which diverse formal and informal information flows and procedures intersect and evolve. Within the third component, a more standard infrastructure provides a substrate for a variety of grids and for archive.

Choo (1995) discusses three different forms of knowledge: tacit knowledge, background knowledge and rule-based knowledge. All three forms may be present on any occasion, but let's consider briefly some of the knowledge type mappings for the three components of Fig. 3. Tacit or implicit knowledge grounded in action or tasks dominates along with information

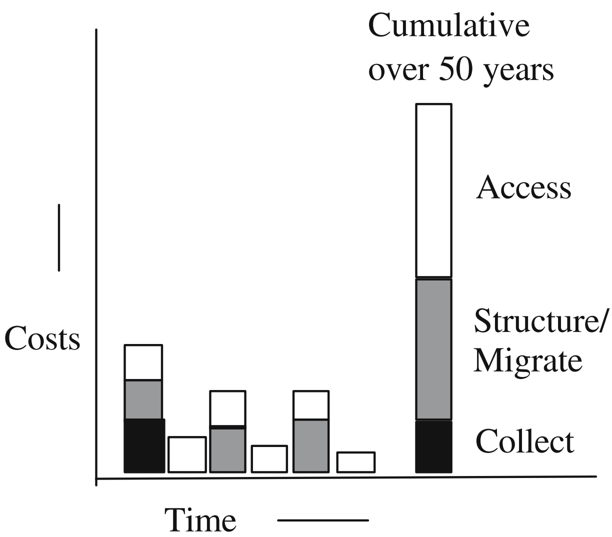

Fig. 2 An indication of costs for 'data care' over time (after Brand, 1994) with a cumulative summation over 50 years shown to the right. *Black* indicates costs of collection; *grey* indicates costs to structure or migrate data; *white* indicates costs to maintain access to data

flows in a project or field area; background knowledge communicated by stories and texts is prevalent in the community arena; and rule-based knowledge provides a foundation for the networked arena. All these, including a fourth class of knowledge, meta-knowledge that integrates across intellectual resources, are subjects of ongoing study.

Of course the process of 'knowledge making' occurs within each of the model components, but it is shown at the center stage as the dominant product. Focusing on 'knowledge making' serves as a reminder of the need for a range of types of information systems and infrastructures to support learning, from ephemeral prototype to production systems and social networks. Information systems that are both functional and used need be a blend of technical and social, that is, a mix of sociotechnical elements. Drawing upon Star and Bowker's (2002) notion of 'to infrastructure', which presents infrastructure as an ongoing activity, the dominant process at the center stage of the information ecology is labeled 'infrastructuring.' The dynamic nature of this center stage is emphasized by the two arrows emphasizing the iterative nature of knowledge making and of learning. The flow of data to global networks depends upon another mix both of physical aspects such as Internet bandwidth and remote delivery mechanisms and of sociotechnical arrangements such as grids and portals.

Characteristics of such a three-component information ecology are gathered in Table 2. Here the first component is labeled 'project' drawing upon a traditional disciplinary field or laboratory work unit that collects data, makes a variety of measurements within a defined timeframe. The work is a result of hypothesizing and creates additional hypothesizing as well as produces papers. The second component, labeled 'community', involves local work with the data where sociotechnical dynamics are a part of ongoing design work and eliciting tacit knowledge, where integration results from the balancing of tensions and the weighing of options. This work overflows with small and large shared activities, each of which may be considered a 'shared boundary object' (Star & Griesemer, 1989). These activities ground the work, create interfaces of understanding, and provide opportunities for mutual learning. The third component, labeled 'partner', focuses on the distribution of information into the global realm, requiring a standardized organization with technical focus on connections and explicit knowledge. It's a federating center bringing together multidisciplinary partners through portals. The fluidity of today's connections readily brings to mind examples that cannot be neatly categorized or classified. For instance, where does the federated project or the heterogeneous partnership fit? Table 2 is meant to be suggestive of the many complex parts of an information ecology, not to represent a definitive list.

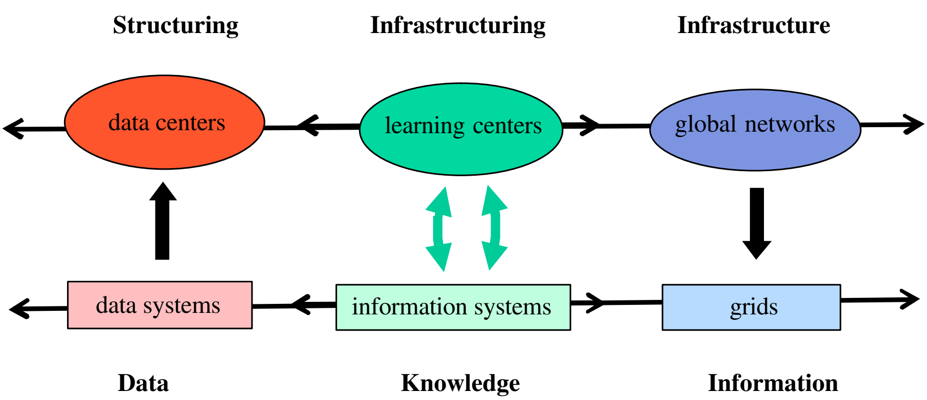

Fig. 3 Information ecology: three component conceptual framework with principle element of data centers, learning centers, and global networks characterized as data, knowledge, and information, respectively

5.2 Information ecology—information manager role

The position of the intermediary is always very hard to pin down—whether the intermediary in question is a person or a program. One of the fuzziest and yet most important concepts in computer science in the past decade has been that of ‘middleware’—that which stands between two programs and allows them to interoperate. Taking the ‘middleware’ and ‘mediation’ concepts into the field of Library and Information Science, for example, it has been very hard to define the creative and innovative work that either librarians or information managers do—even though many attempts at ‘disintermediation’ have failed. Software engineers write programs that can be demonstrated in conferences and written up in journals. Domain scientists produce data that can be run through a research protocol and published in a journal. Information managers on the other hand service, manage, and design the flow of information (as do librarians). They take the materials—organizational, technical and data—which are at hand and make it all work together. Their work is rarely written about; when spoken of, it frequently has the ‘what I did during my holiday’ patina: it is too specific to generalize and seems too small scale to label important. It is work of bricolage as much as work of engineering, in Levi-Strauss’s (1966) terms.

Emerging understandings suggest a need to (re) consider tasks and roles. Domain experts may not separate the management of information from the management of technology. Information experts may not articulate the variety of information processes including identifying information needs, acquiring information, organizing diverse sources, storing materials, developing information products and services, distributing information, and using information for analysis. The fields of engineering and design requirements, usability and record keeping, are contributing to new models, broadening information management beyond a restricted, reactive, service role but their use and application in ongoing ecological projects requires currently unsupported scaling of design and communication.

The work of being an information manager, frequently calling for creativity and innovation, is most often an integral part of an information flow. We just don’t have good ways of talking about the work or measures of its success: the creative aspects of the work of the information manager get buried under its image as purely a service occupation. The Science Studies and CSCW communities are contributing to the visibility of this work with a vocabulary for everyday practices, such as tacit knowledge, workarounds, alignments, articulation work, and intermediation. In Fig. 3, we characterize this as the work of ‘infrastructuring.’ It is work, which stands between the work of the domain scientist on the one hand, and the computer scientist on the

Table 2 Information ecology components

Project	Community	Partner
field	local	global
data collection	knowledge making	information flow
team	community-of-practice	organization
practice	sociotechnical	technical
milestones	designs	connections
measurement	tacit	explicit
heterogeneity	integrative	homogenous
hypothesizing	learning	federating
#papers	#boundary objects	#partners
#products	#processes	#portals
disciplinary	interdisciplinary	transdisciplinary

other. Its object stands between the databank on the one hand and the global network on the other—the goal being to fit these two to each other. This ‘in between’ work does not have fixed boundaries, protocols or procedures—indeed it is defined by the lack of them.

To create opportunities for mutual learning and iterative feedback, the sharing of observations through ongoing dialogue is one goal of our participatory design study. Evidence of unarticulated informal knowledge emerges from our LTER ethnographic study. Tensions evident in handling data were articulated in interviews with information managers and summarized in a diagram presented to the community in 2002 (Fig. 4a). Subsequent analysis resulted in a re-representation of the complex information management tasks involving local knowledge and everyday experience into a triangle of interdependent elements. The LTER information manager work is shown unpacked into three distinct tasks: service, management and design (Fig. 4b; Karasti & Baker, 2004).

So why should we be paying more attention to this liaison work, this work of intermediation? The traditional ”progressive” view of infrastructure development conceives the ideal process as one of disintermediation—to present the scientist with a transparent window on the data through the window of the technology. In this process, not only does the work of the information manager get dropped out of the equation (just as a catalyst from a chemical reaction) but also the end-user gets represented as an individual sitting in front of a screen (Fig. 5a).

What is missing here is precisely the centrality of the ongoing work of intermediation that brings users and information managers together into the center of the mix—with ‘science,’ ‘data’ and ‘technology’ as contingent outcomes of the intermediation process. In contrast, Fig. 5b moves both liaison roles and the users to a location more central to science-data-technology activities. Such a move creates a system with participants central to design and positioned to give feedback for ongoing system redesign.

6 Conclusion

Data appearing in databases is a partial representation of our understanding of our world. This understanding is held in place by a matrix of organizational concerns, policy judgments and

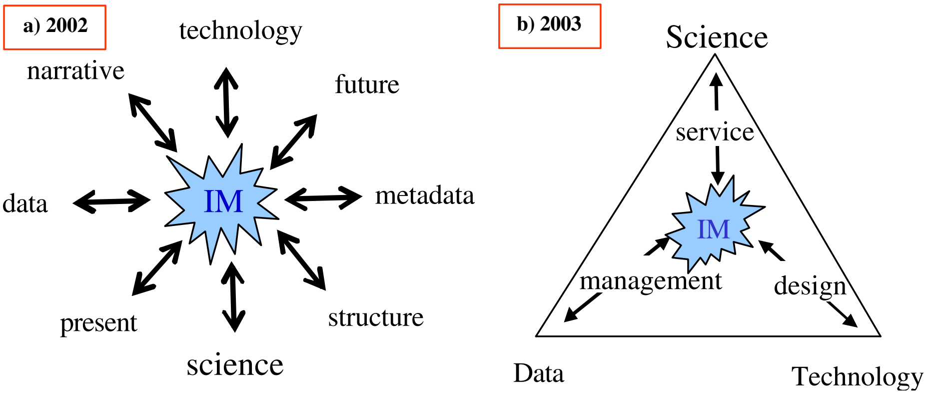

Fig. 4 Presenting LTER Information Manager voices: **a** tensions associated with information management. **b** representation of information management in a mediating position in relation to science, data, and technology with respect to three distinct roles, service, management, and design

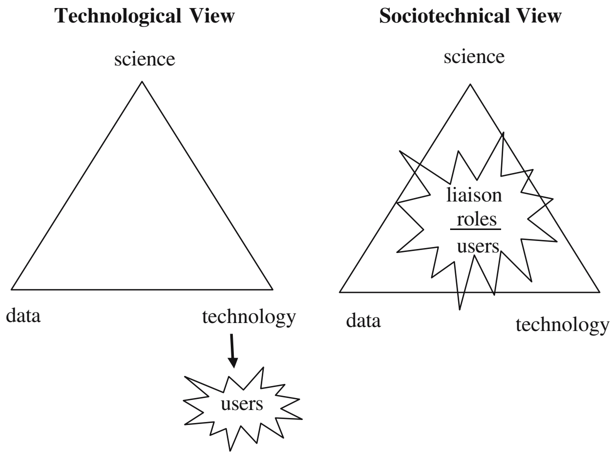

Fig. 5 Presenting two perspectives on the relationship of users with respect to the realm of science-data-technology. The technological view (*disintermediated*) takes into account the user as an 'outside' entity contrasts with the sociotechnical view (*intermediated*) with the users as well as liaison roles appearing centrally

scientific practices. Data in the database are the result of a multitude of negotiated processes from sampling design choices to data collection methodologies, from calibration issues to quality assessments, from analysis algorithms to data presentations, from conceptual mappings to knowledge synthesis. From the diverse flows of information, forms of knowledge, and interrelationships between them, the view of an information ecology as an open system arises. Tensions are a necessary formal feature of the system, requiring (re) balancing as our understandings deepen and our views broaden. Tensions and balances may be explored through ethnographic studies and by varying the unit of study, perhaps from the data to memory practices or the research process. Complexity, ambiguity and nonlinearity are a part of an information ecology and may be addressed today by considering multiple types of knowledge, developing information system vocabularies, and recognizing the emergent need for intermediation.

The conceptual framework for the information ecology presented offers a process oriented approach to information systems. Process oriented work is frequently invisible and rarely supported, so traditionally it has been left undone or only sporadically addressed. In foregrounding the process itself, invisible work is identified. Interdisciplinary teams comprised of domain, information and social scientists can help give this invisible, articulation work the analytic status it deserves. As recent NSF reports on cyberinfrastructure note (Atkins & NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure, 2003; Futrell & AC-ERE, 2003; Pfirman, 2003), both computer and information mediation are central to the practice of science in the twenty-first century. In participating in interactive design, sharing boundary objects, and building common vocabularies, knowledge making is facilitated across its multiple dimensions. Alongside the work of domain ontology building and national standards development, we need to articulate and model the continuing work of intermediation that perpetually modifies ontologies and standards according to local contingencies.

From attention to informal knowledge and long-term care of data diversity (Bowker, 2000), comes the need to consider data stewardship and information system openness. This tension between homogenization and diversification is a part of a larger tension between technological determinism (the killer app will fix it) and methodological relativism (all data is ineluctably local). The question then is not only 'with which epistemological and ontological frameworks shall we work?' but also 'how can we work at the intersection between different frameworks?' With the need for interdisciplinarity and interoperability come not fewer but rather added dimensions to balance and boundaries to walk.

In building information systems to support ongoing work and databases for both current and future use, we must consider the dilemmas that representation of an earth system poses inherently, such as which data to capture and how to capture that data's context; how to capture informal knowledge and what memories to preserve. Scaling issues for a digital world provide opportunities for insight into new approaches and a (re)viewing of the ecosystem issues faced in the natural world. Slipping Brand's architectural insights (1994) explicitly into an information systems context: An information environment isn't something you finish; it's something you start. This refocuses information management work within an information ecology on "How Databases Learn." Since change won't go away, the question is not just 'how do we capture knowledge?' or 'how do we deal with uncertainty and change in an information system?' but rather 'how do we build an open information environment?'—a purposefully leaky, data-diverse, tension-balancing information ecology that recognizes uncertainty and facilitates change.

Acknowledgements The quotes for this paper are from transcriptions of interviews conducted by Helena Karasti. We offer special recognition to this member of our research team. We thank the LTER community members. NSF Grants EIA-01-31958, SBE/SES & OPP-04-05069, and OPP-02-17282 as well as Scripps Institution of Oceanography for support of this work.

References

Ashby, W. R. (1956). *Introduction to cybernetics*. London: Chapman & Hall.

Atkins, D., & NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure. (2003). NSF-AP Report: Revolutionizing Science and Engineering Through Cyberinfrastructure.

Babbage, C. (1837). *The ninth bridgewater treatise, a fragment*. London: John Murray.

Baker, K. S., Benson, B. J., Henshaw, D. L., Blodgett, D., Porter, J. H., & Stafford, S. G. (2000). Evolution of a multisite network information system: The LTER information management paradigm. *BioScience, 50–11*, 963–978.

Baker, K. S., Bowker, G., & Karasti, H. (2002). Designing an infrastructure for heterogeneity in ecosystem data, collaborators, and organizations, in *Proceedings of the second national conference on digital government research* (pp. 141–144). Los Angeles, CA. http://www.dgrc.org/dgrc/dgo2002/.

Boland, R. J., & Tenkasi, R. V. (1995). Perspective making and perspective taking in communities of knowing. *Organization Science, 6–4*, 350–372.

Bowker, G. C. (2000). Biodiversity, datadiversity. *Social Studies of Science, 30–5*, 643–684.

Bowker, G. C. (2006). *Memory practices in the sciences*. Cambridge, MA: MIT.

Bowser, C. J. (1986). Historic data sets: Lessons from the past, lessons for the future. In W. T. Michener (Ed.), *Research data management in the ecological sciences* (pp. 155–179). Columbia, South Carolina: University of South Carolina Press.

Brand, S. (1994). *How buildings learn*. New York: Penguin.

Brunt, J. W. (1998). The LTER network information system: A framework for ecological information management. In *Proceedings (RMRS-P-12) of North American science symposium—towards a unified framework for forest ecosystem monitoring and research* (pp. 435–440). Guadalajara, Jalisco, Mexico.

Chalmers, A. (1976). *What is this thing called science: An assessment of the nature and status of science and its methods*. Cambridge: Hackett.

Choo, C. W. (1995). Information management for the intelligent organization: Roles and implications for the information professions. In Digital Libraries Conference.

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

Deetz, S. (1996). Describing difference in approaches to organization science; rethinking burrell and morgan and their legacy. *Organization Science, 7–2*, 191–207.

Douglas, M. (1986). *How institutions think*. Syracuse, NY: Syracuse University Press.

Eaton, A. J. J., & Bawden, A. D. (1991). What kind of resource is information. *International Journal of Information Management, 11*, 156–165.

Ecological Visions Committee (2004). *Ecological science and sustainability for a crowded planet*. Ecological Society of America.

Edwards, T. C., Homer, C. H., Bassettt, S. D., Falconer, A., Ramsey, R. D., & Wight, D. W. (1995). *Utah GAP analysis: An environmental information system*. *Technical Report 95–1*. Logan, Utah: Utah Cooperative Fish & Wildlife Research Unit, Utah State University.

Eriksen, T. H. (2001). *Tyranny of the moment: Fast and slow time in the information age*. London: Pluto.

Finholt, T. (2002). Collaboratories. In E. B.Cronin (Ed.), *Annual review of information science and technology, vol. 36* (pp. 73–107). Medford, NJ.

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

Futrell, J., & AC-ERE. (2003). Environmental cyberInfrastructure: Tools for the study of complex environmental systems AC-ERE. http://www.nsf.gov/ere.

Gasson, S. (2004). The management of distributed organizational knowledge. In Proceedings of the Hawaii international conference on information systems.

Gosz, J. (1999). International long term ecological research: Collaboration among national networks of research sites for a global understanding, long term ecological research: Examples, methods, perspectives for Central Europe, Madralin, Poland, International Centre of Ecology, Polish Academy of Sciences.

Greenland, D., Goodin, D. G., & Smith, R. C. (2003). An introduction to climate variability and ecosystem response. In D. Greenland, D. G. Goodin, & R. C. Smith (Eds.), *Climate variability and ecosystem response at long-term ecological research sites* (pp. 3–19). New York: Oxford University Press.

Henshaw, D. L., Stubbs, M., Benson, B. J., Baker, K. S., Blodgett, D., & Porter, J. H. (1998). Climate database project: A strategy for improving information access across research sites. In W. K. Michener, J. H. Porter, & S. G. Stafford (Eds.), *Data and information management in the ecological sciences: A resource guide* (Proceedings of workshop, held at University of New Mexico, Albuquerque NM, 8–9 August, 1997) (pp. 123–127). Albuquerque, NM: Long-Term Ecological Research Network Office, University of New Mexico.

Hobbie, J. E. (2003). Scientific accomplishments of the long-term ecological research program: An introduction. *BioScience, 53*, 17–20.

Hughes, T. P. (1983). *Networks of power: Electrification in western society, 1880–1930*. Baltimore, MD: John Hopkins University Press.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT.

Iivari, J. (1991). Paradigmatic analysis of contemporary schools of IS development. *European Journal of Information Systems, 1–4*, 49–272.

Jordan, B. (1996). Ethnographic workplace studies and CSCW. In D. Shapiro, M. Tauber, & R. Traunmuller (Eds.), *The design of computer supported cooperative work and groupware systems* (pp. 17–42). Amsterdam: Elsevier.

Kaiser, J. (2001). An experiment for all seasons. *Science, 293–5530*, 624–627.

Kaplan, S., & Seebeck, L. (2001). Harnessing complexity in CSCW. In Proceedings of the seventh european conference on computer supported cooperative work (pp. 359–397), Kluwer.

Karasti, H. (1994). What's different in gender oriented ISD? Identifying gender oriented information systems development approach. In E. A. Adam (Ed.), *Women, work and computerization* (pp. 45–57). North-Holland: Elsevier.

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

Kinzig, A. P., Carpenter, S., Dove, M., Michael, M., Heal, G., Levin, S., et al. (2000). Nature and society: An imperative for integrated environmental research. In Executive summary of a workshop sponsored by NSF, Developing a Research Agenda for Linking Biogeophysical and Socioeconomic Systems (p. 72). Tempe, Arizona. http://lsweb.la.asu.edu/akinzig/report.htm.

Levi-Strauss, C. (1966). *The savage mind*. London: Weidenfeld and Nicolson.

Likens, G. E. (1989). *Long-term studies in ecology: Approaches and alternatives*. Berlin Heidelberg New York: Springer.

Magnuson, J. J. (1990). Long-term ecological research and the invisible present—Uncovering the processes hidden because they occur slowly or because effects lag years behind causes. *BioScience, 40–7*, 495–501.

Manovich, L. (1999). Database as a symbolic form. *Millenium Film Journal, 34* (Fall).

Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B., & Stafford, S. G. (1997). Nongeospatial metadata for the ecological sciences. *Ecological Applications, 7–1*, 330–342.

Michener, W. K., Brunt, J. W., & Stafford, S. G. (1994). *Environmental information management and analysis: Ecosystem to global scales*. London: Taylor & Francis.

NRC. (2001). *Grand challenges in environmental sciences*. Washington, DC: National Academy Press.

Odum, E. P. (1995). The emergence of ecology as a new integrative discipline. *Science, 195*, 1289–1293.

Odum, E. P. (1998). *Ecological vignettes: Ecological approaches to dealing with human predicaments*. Amsterdam: Harwood Academic Publishers.

Ohman, M. D., & Venrick, E. L. (2003). CalCOFI in a changing ocean. *Oceanography, 16*, 76–85.

Orr, D. W. (2002). *The nature of design: Ecology, culture, and human intention*. Oxford: Oxford University Press.

Pfirman, S. (2003). Complex environmental systems; synthesis for earth, life and society in the 21st Century. A report summarizing a 10-year outlook for the National Science Foundation.

Pickett, S. T. A., & Cadenasso, M. L. (2002). The ecosystem as a multidimensional concept: Meaning, model and metaphor. *Ecosystems, 5*, 1–10.

Poore, B. (2003). Blue Lines: Water, Information, and Salmon in the Pacific Northwest. Ph.D Thesis, University of Washington. 335p.

Redman, C., Grove, J. M., & Kuby, L. (2000). Toward a unified understanding of human ecosystems: Integating social sciences into long-term ecological research. In White Paper of the Social Science Committee of the LTER Network. http://www.lternet.edu/documents/Publications/sosciwhtppr/index.html.

Robertson, P. D., Coleman, C., Bledsoe, C. S., & Sollins, P. (1999). *Standard oil methods for long-term ecological research. Long-term ecological research network series*. New York: Oxford University Press.

Serres, M. (1990). *Le Contrat Naturel*. Paris: F. Bourin.

Smith, M. R., & Marx, E. L. (1994). *Does technology drive history? The dilemma of technological determinism*. Cambridge: MIT.

Solomon, P. (1997). Discovering information behavior in sense making. I. Time and timing. *Journal of the American Society for Information Science, 48–2*, 1097–1108.

Spasser, M. A. (1997). The enacted fate of undiscovered public knowledge. *Journal of the American Society for Information Science, 48–8*, 707–717.

Star, S. L., & Bowker, G. C. (2002). How to infrastructure. In L. A. Lievrouw & S. L. Livingstone (Eds.), *The handbook of new media* (pp. 151–162). London: SAGE.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, “translations,” and boundary objects: Amateurs and professionals in Berkeley’s museum of vertebrate zoology, 1907–39. *Social Studies of Science, 19*, 387–420.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information systems. *Information Systems Research, 7–1*, 111–134.

Star, S. L., & Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *CSCW, 8*, 9–30.

Suchman, L. (2000). Organizing alignment: A case of bridge-building. *Organization, 17–2*, 311–327.

US JGOFS (2001). Ocean biogeochemistry and the global carbon cycle: An introduction to the U.S. joint global ocean flux study. *Oceanography, 14–4*, 5–121.

Walsh, J. P., & Ungson, G. R. (1991). Organizational memory. *Academy of Management Review, 16–1*, 57–91.

Wauzzinski, R. A. (2001). *Discerning prometheus: The cry for wisdom in our technological society.* Madison: Associated University Press.

Weick, K. E., & Sutcliffe, K. M. (2001). *Managing the unexpected, assuring high performance in an age of complexity.* San Francisco, CA: Jossey-Bass.

Weick, K., Sutcliffe, K., & Obstfeld, D. (2005). Organizing and the process of Sensemaking. *Organization Science, 16–4,* 409–421.

Whitley, E. A. (2000). Tacit and explicit knowledge: Conceptual confusion around the commodification of knowledge. In Conference proceedings of knowledge management: Concepts and controversies (pp. 62–64). University of Warwick.

Yates-Mercer, P., & Bawden, D. (2002). Managing the paradox: The valuation of knowledge and knowledge management. *Journal of Information Science, 28–1*, 19–29.

Zimmerman, A. S. (2003). *Data sharing and secondary use of scientific data: Experiences of ecologists*. PhD Thesis. Ann Arbor: The University of Michigan.