Scientific collaboration and information infrastructures – Information practices in LIDET-experiment

Eija Halkola

Department of Information Processing Sciences, University of Oulu, P.O. Box 3000, FIN-90014 University of Oulu, Finland

halkola@paju.oulu.fi

***Abstract**: This paper studies scientific collaborations and collaborative infrastructures through focusing on everyday information practices of data management of the distributed ecological Long-Term Intersite Decomposition Team (LIDET)-experiment. LIDET, being part of the Long Term Ecological Research (LTER) network, has placed an emphasis on long-term preservation and sharing of data. The research is empirically grounded and applies qualitative research approaches and a socio-technical perspective aiming to produce descriptive understandings of the aspects of information practices including planning, implementation and development of information infrastructures in the LIDET-case study. Collaborative preplanning and centralized data management created a* *sustainable information infrastructure. The infrastructure was also able to adjust to changes in working practices, an increase in the number of collaborators and developments in technology. Empirical material used in this case study is collected through ethnographic fieldwork in the Long-Term Ecological Research (LTER) network during the year 2002.*

Keywords*: collaborative infrastructure; information practices, use and long term preservation of data; metadata; distributed scientific experiments; organization; coordination; practical experience*

1. Introduction

The research relates to the context and studies of scientific collaborations (Karasti et al., 2004) and collaborative infrastructures (Star & Bowker, 2002; Star & Ruhleder, 1996; Baker et al. 2002) by studying the information practices of a networked ecological experiment and its infrastructural support for collaboration. In this case study of the LIDET-experiment, in collaboration with Long Term Ecological Research (LTER) network, the emphasis is on the problematics of distributed scientific collaboration and long-term information management. Research is empirically grounded and applies a socio-technical perspective (Kling 1999) aiming to produce descriptive understandings of the elements of information practices and collaboration infrastructures of the case organization. The main information practices in LIDET-experiment relate to gathering and storing environmental data, creating metadata and designing and maintaining databases.

The focus in this case study has been on the *Community of Practice* (CoP) (Lave & Wenger, 1991) that formed the research coordination and data management group called "Litterbag Central" at the H.J. Andrews site, Oregon State University (OSU). Communities of practice are groups of people who share similar goals and interests. In pursuit of these goals and interests, they employ common practices, work with the same tools and express themselves in a common language. The CoP, data management group of LIDET consisted of the information manager of the coordinating site, the principal investigator, two research assistants, computer support employees and temporary assistants. For information manager, the data manager group consisting of a representative from each LTER site, also provided a forum in its annual meetings for cross-site conversations, collaborative project development and joint network technology strategizing (Karasti & Baker 2004).

In LIDET, as in the LTER context in general, the information managers have had a fundamental role in both developing the technological infrastructure (Baker et al. 2000) and in providing ongoing support for science and data care (Karasti & Baker 2004). Ecologists of the LIDET-team as well as other LTER scientists, responsible for collecting and interpreting data, benefit from working in partnership with data and information managers, to ensure optimal data and information availability (Baker et al., 2000).

Recent advances in communication technologies and infrastructure have brought new possibilities for scientific collaboration and networking. Science in general increasingly depends on geographically distributed collaboration. (Finholt 2002.) Knowledge is constructed jointly at a distance by multidisciplinary teams supported by electronic communication and information infrastructures (Kanfer et al., 2000). There exists a discrepancy between the potential to understand the contextual knowledge practices and employ the situated knowledge of scientific collaborations, and the limited ways in which the current information infrastructures can support these kinds of knowledge practices in the context of extended collaborations (Kanfer et al. 2000).

"Embedded" authentic knowledge of actors is entrenched within specific Communities of Practice (Lave and Wenger, 1991). Through analysing the empirical research material of this case study by applying a socio-technical approach (Kling 1999) and qualitative analysis methods, it is possible to study the data management of the LIDET-experiment embedded within planning, organizing and coordinating the ecological research. Some of the situated knowledge related to everyday data management practices of LIDET are also described and discussed. Focusing on everyday information practices is motivated by profound theoretical notions of working practices (Karasti 2001, cf. Karasti and Baker 2004). Information practices in this case study context relate to practices of information gathering, archiving, retrieval, use, and sharing.

In this study information infrastructures are seen as socio-technical constructs (Star & Ruhleder 1996, Star & Bowker 2002) that operate in specific social and organizational settings. This paper focuses upon the notions of infrastructure suggested by Star and Ruhleder (1996), which are used as a framework to describe the information practices in creating infrastructure for scientific collaboration. Star and Ruhleder(1996) see infrastructure as part of human organization. In addition Star and Bowker (2002) define infrastructure as being relative to working conditions and argue that it never stands apart from the people who design, maintain and use it. By foregrounding the background elements of information practices of the case, this case study performs Bowker's (1994) concept of infrastructural inversion.

As background information this paper describes the context of the target organisation and explains research materials and methods used in this case study. The elements of data management of the long-term ecological LIDET-experiment are described including the intertwined preplanning of research and data management and also data management during the long-term research.

This paper also presents the characteristic aspects, distributedness and long-term perspective of the LIDET-experiment, which strongly defined the data management practices and challenges addressed to its information management. The centralized research and data management as well as the adaptation and modification of research and data management practices are presented as the main solutions to these data management challenges.

2. Background

Complex issues (e.g., global change, sustainability, biodiversity, and emerging diseases) require interdisciplinary collaboration and synthesis over much broader spatial and longer temporal scales (Levin, 1992; Kareiva & Anderson, 1988).

The Long-Term Ecological Research (LTER) program was initiated in 1980 by the National Science Foundation (NSF). The central, organizing intellectual aim of the LTER program is to understand long-term patterns and processes of ecological systems at multiple spatial scales (Hobbie et al., 2003). Today LTER is a federation of twenty-four independent research sites, one network office site in the United States (Hobbie et al., 2003) and in addition a developing International LTER (Gosz, 1999). The LTER Network initial vision (Franklin et al., 1990) and continuing mission (Hobbie et al., 2003) include the concept of data management, requiring it to be part of each research site's science plan. Since the mid 1990's, each site is required to have data available on the Internet two years after its collection. In information management LTER is faced with the specific challenge of how to maintain datasets over the long-term (Michener et al., 2000) which are useable and understandable after decades, possibly even after centuries (Baker et al., 2002).

2.1 Empirical background

LIDET-experiment (http://www.fsl.orst.edu/lter/research/intersite/lidet.htm), in collaboration with LTER network, was a 10-year, inter-site research to test the effect of substrate quality and macroclimate on long-term decomposition and nutrient dynamics of fine litter. The LIDET decomposition experiments have been installed at distributed sites that span a wide array of ecosystems including the 17 LTER sites and 4 non-LTER sites. Seven additional sites were added in 1991 bringing the total to 28 participating sites. Team experiments, such as LIDET, designed and set up as an intersite experiment and carried out simultaneously at many sites may help us to understand ecosystem behavior over greater temporal and spatial scales. (Long-Term Intersite Decomposition Team, 1995.) The LIDET-experiment was coordinated inside the LTER-network and benefited from the existing technical infrastructure developed within the LTER-network.

The research material flow was centrally organised and coordinated in the laboratory of the H.J. Andrews LTER site in order to minimize the field collaborators' responsibilities. The Data management group from the organising site was responsible for chemical analysis, data management and preliminary data analysis (Long-Term Intersite Decomposition Team, 1995). During the experiment, data management personnel became experts of long-term research and were asked for advice related to research supplies and arrangements. Samples were weighed, handled and analyzed in one laboratory with standardized methods, thus assuring the comparability of the results of chemical analyses. Standardized methods made the results comparable so that individual sites were able to place their site-specific research results in a larger context and benefit from the research approach. The individual sites also benefited from general access to the Near Infrared Chromatograph (NIR) analytical method.

Data and credit for the LIDET-experiment were shared according to agreed guidelines. Each site annually received a current, proofed copy of the data from the site. Site investigators had one year in which to prepare site-specific manuscripts, to be published under individual names. After one year, the data became available for intersite syntheses to be published under joint LIDET authorship and for model parametrization and testing. (Long-Term Intersite Decomposition Team, 1995.)

The interview data outlined the importance of data management preplanning of the long-term preservation of data. Preplanning of research procedures and data management of the LIDET-experiment included preplanning of everyday working practices as well as annual preplanning. A member of the LIDET personnel describes the preplanning and inter-site nature of the experiment:

"The difference with LIDET was that LIDET was a preplanned … the whole thing was designed to do an inter-site work… It was planned out to do things with exactly the same methods across multiple sites, and doing it with a random statistical design so they could actually have statistical hypothesis and results." (LIDET personnel)

Collaborative planning of data management and technology development was closely intertwined with planning, organisation and coordination of research procedures and information practices, aiming to provide support for science. The contribution of the experienced information management to planning was to ensure the preservation of the collected data in long-term databases. Careful planning also saved resources in the data-handling phase.

Two research assistants were mainly responsible for the everyday data management of the LIDET-experiment. Study packages as well as instructions for sample collection were prepared and sent to all field collaborators annually. Instead of sending sites only detailed instructions of how to implement the experiment, research assistants sent very carefully designed and pre-planned research packages, ensuring the unity of the used methods and the carrying out of the experiments. Research sample numbers were generated with the LIDET-program from the LIDET-database and associated with the sample data sheets and labels.

Each collaborating site received ten types of leaf and root litter and also wooden dowels. Litterbags were installed once at all 28 collaborating sites and collected annually during the experiment. Collaborators at the individual sites were responsible for collecting the litter samples, oven-drying, weighing and sending samples to OSU for chemical analysis, data entry and long-term storage. (Long-Term Intersite Decomposition Team, 1995.)

Data management during the research procedure was time consuming since it required a lot of organization and coordination of the research material flow. Research assistants received the collected, oven-dried and weighed samples with datasheets from sites. All received samples were tagged, weighed and data was checked before being sent back to sites for verification.

LIDET- computer software program was developed for managing all research data. Data management personnel were responsible for entering data in the LIDET-database and also for quality control and assurance of data. Controlled identification numbering of each sample benefited in the detecting of anomalies or unexpected measurements and quality control. The LIDET database was eventually stored in the OSU Forest Science Data Bank (FSDB), developed and maintained by the H.J. Andrews site to store the LTER and other ecological data sets (Stafford et al., 1988).

2.2 Theoretical background

Studying infrastructures in the LIDET-experiment context means a focus on information practices in data- and information management work and in developing and maintaining infrastructures for scientific collaboration. Many aspects of infrastructure are "singularly unexciting" suggesting that the study of infrastructure work is "studying boring things" (cf. Star, 1999). In the LIDET case this includes investigating the gradual process of the modification of working practices and maintenance of infrastructures (Star & Bowker 2002), for instance the sewing and packing of sample bags.

Star and Ruhleder's (1996) notion of infrastructure is grounded in a perspective suggested by Bowker and Star (1999) where information infrastructures are perceived as performative, that is, partially creating the world they subtend and the tradition of science and technology studies. According to Star and Ruhleder (1996) infrastructure is a fundamentally relational concept; infrastructure becomes real in relation to organized practices. Star and Ruhleder (1996) define the qualities and salient features of infrastructure:

- *Embeddedness*. Infrastructures are sunk into, inside of, other structures, social arrangements and technologies.

- *Transparency*. Infrastructure is transparent to use, in the sense that it does not have to be reinvented each time or assembled for each task, but invisibly supports those tasks.
- *Reach or scope*. This may be either spatial or temporal: infrastructure has reach beyond a single event or one-site practice.
- *Learned as part of membership*. The taken-for-grantedness of artifacts and organisational arrangements is learned as part of membership in a community of practice (Lave and Wenger, 1992).
- *Links with conventions of practice*. Infrastructure both shapes and is shaped by the conventions of a community of practice.
- *Embodiment of standards*. Modified by scope and often by conflicting conventions, infrastructure takes on transparency by plugging into other infrastructures and tools in a standardized fashion.
- *Builds on an installed base*. Infrastructure does not grow *de novo*; it wrestles with the "inertia of the installed base", and inherits strengths and limitations from that base.
- *Becomes visible upon breakdown*. The normally invisible infrastructure becomes visible when it breaks.

This paper focuses on the information practices of distributed and long-term ecological research and its infrastructural support for collaboration, including the development and maintenance of information infrastructures.

2.3 Research method

The Field of Science and Technology Studies applies ethnomethodological studies to provide understanding of work practices in scientific work at the concrete level (e.g. Star & Ruhleder 1996; Bowker et al. 1997). This study focuses on scientific collaboration and information practices to produce descriptive understandings of the elements of information practices and collaboration infrastructure of the case.

This research analyses the ethnographic materials collected by professor Helena Karasti within LTER network (http://www.lternet.edu/) in a BioDiversity and EcoInformatics (BDEI) project in 2002. The main empirical material for analyses consisted of interview data of LIDET participants. In addition, research material included scientific articles of the LTER-network and related literature of information management, extensive LTER Internet pages and digital pictures. In the qualitative content analysis of interview materials my aim has been to capture different views of the experiment from within the community. By applying a socio-technical approach and qualitative content analysis this case study also outlines the social aspects in the analysis of the reserach material.

Qualitative content analysis, being most usable for analysis of textual data (Catanzaro 1988, Krippendorff 1981), was carried out as descriptive inductive data analysis of the research data, where created codes were based on the research material (Catanzaro 1988). Content analysis is based on interpretation and induction in a process which proceeds from the empirical material to a more conceptualized analysis (Tuomi and Sarajärvi 2002).

The main categories based on the content analysis were related to information and data management during different phases of research, and to categories describing distributedness and long-term perspective of the LIDET experiment. The notions of infrastructure (Star & Ruhleder 1996) were related to the information practices of LIDET-experiment in order to explore the elements of long-term collaboration infrastructure and also increase sensitivity towards social issues of networked ecological research and its infrastructural support.

The qualitative analysis of interview material was started individually by the author during year 2003. The excerpts from the interview materials have also been discussed in collaborative data analysis sessions. The excerpts of interviews in this paper are marked with acronyms.

3. Long-term and distributed scientific collaboration

The LIDET- experiment was a widely distributed, centrally coordinated and loosely coupled, 10 year ecological research experiment where collaboration was organized indirectly through the research material specimens and data. The characteristic aspects of LIDET scientific collaboration, i.e. distributedness and long-term perspective, define its central information practices and collaboration infrastructures. These characteristic aspects of the LIDET-experiment addressed challenges to data management and also to the maintenance of the infrastructure.

3.1 Long-term perspective

Long-term science is concerned with the research need to collect and archive data over long time periods (Karasti & Baker 2004). The interview material emphasized that planning the data management for the long-term and distributed research collaboration must be based on the preplanning of the experimental design and also anticipation of possible problems during the long-term in order to enable the successful carrying out of the long-term experiment.

Designing infrastructures and making decisions about new technologies addressed a concern of how to minimize the risks for data archival for long-term data management and science support. New pieces of technology in LTER are evaluated against their value for ecological research. Before being embedded into existing infrastructures, new technologies also have to be aligned with existing information practices and infrastructure. (Karasti & Baker 2004.)

3.1.1. Work practice development

The need for improvements in working practices became obvious in different phases of the long-term LIDET-experiment. According to Star and Ruhleder (1996) infrastructure is shaped by developing work practices and changing conventions of a community of practice. In the centrally managed LIDET-experiment embedded knowledge of data management personnel could be applied to modifying the infrastructure in order to support scientific collaboration. During the experiment artifacts were developed to make the research preparations and coordination more flexible. The developed research artifact, named as a pneumatic stapler, allows assistants to close a sample bag with three staples at a time, making the sample bag preparations faster. To describe the everyday innovation, a member of the LIDET personnel tells "an artifact story" of the pneumatic stapler developed especially for this experiment:

"OK, LIDET has made, has made us the litter decomposition experts so everybody is contacting us all the time for stuff and they're always wanting to borrow… this multi-head stapler! For it's pneumatically operated, … Because nobody sells anything like this and so people email us saying how frustrated they are trying to find something that could do this and could they please, please borrow yours, and so this spends a lot of time travelling. It just came back from Yellowstone and tomorrow it's going to British Columbia."(LBC)

Modifications to working practices were also needed because of un-anticipated field situations during sample collecting. Sample bags of the same set were sewn on a string and each string was marked with an identification number. This everyday work practice innovation of tying sample bags together in order to minimize the risk of sample bags getting mixed was based on situated knowledge of data management personnel. This embedded feature of the infrastructure (Star & Ruhleder 1996) became visible during special field circumstances. A member of the LIDET personnel describes research situations in the field:

"… when a fieldworker went out into the field and say an elk had run through and got stuck on the string and dragged it all around … they [field workers] could go and dig around and at least find the tag number so that they would know that they were collecting the right string …" (LBC)

Adapting the experiment arrangements and developing the sample collecting technique made the collaboration in the experiment easier and also assured the success of the experiment.

Data management personnel also created a check list to help the sample collection and data management of the experiment. The check list was used for recording the process of printing the labels and labeling the sample bags and also was needed in order to prevent the losses of research materials. A member of the LIDET personnel describes it this way:

"The next person coming in didn't even know the study was going on, doesn't know the litter bags are, doesn't know what to do with them … so we ended up creating a check list so that every year we could record the process of printing the labels, labelling the bags. " (LBC)

3.1.2. Management of infrastructure changes

Both social and technological infrastructure had to preserve its functionality during a long-term experiment. According to Star and Ruhleder (1996) the normally invisible quality of the working infrastructure becomes visible when it breaks. In LIDET-case infrastructure breaks were caused by technical and social changes. At a social level Star and Bowker (2002) see that flexibility for modifications to information infrastructures enables responding to emergent social needs. A good information infrastructure is stable enough to allow information to be able to persist in time and also be designed for flexibility (Bowker & Star 2002).

The social infrastructure was subjected to changes during a long-term experiment. Changes in research personnel or use of temporary staff causes breaks in social infrastructure, endangering the continuity of data collection and archiving. As an example of a social infrastructure break, a member of the LIDET personnel describes the difficulty faced in delivering a research package:

"… it had happened that the bag had arrived and the person in the front office didn't know who to give it to because the person we addressed it to was no longer there… " (LBC)

In the case of LIDET these personnel changes emphasized the importance of informing collaborators. Annual instructions were sent to field collaborators along with study packages and they were also informed of different stages and changes in research practices during the experiment.

Developing technologies were aligned with existing technologies and practices during the LIDET-experiment. Despite technological development, old technologies need to be maintained long enough to be able to carry out the experiment. The most notable technological change during the LIDET-experiment was the update of the used computer operating system from a DOS-based to windows-based system.

Programs used in LIDET information practices were designed in the DOS-operating system, which caused problems with data management. LIDET was a ten-year experiment and even during that time problems occurred related to the fact that the computer support personnel did not have experience of the earlier DOS-based systems. The change of operating system caused incompatibility problems between originally installed programmes and the new operating system. Normal working practices like generating labels and data sheets failed because the original programs did not function in the new operating system. Knowledge of earlier information practices and infrastructure preserved among computer support staff was important when implementing new technologies. Computer support personnel had to come up with solutions to make systems compatible. A member of the LIDET personnel describes the complicated situation related to implementation of new technologies in this way:

"... we would have to go and talk to our computer support people and they would have to reach back into their memories ... to try and think back and try and think of strange little arcane ways of making the ... computer handle these passé mechanisms... go to a printer capture queue or something ... " (LBC)

In LIDET, technological infrastructure such as old program versions as well as old computers and printers had to be retained in a central laboratory during the experiment to avoid problems in working practices. A new operating system could not be directly applied because conventions of practice were built on an installed base (Star & Ruhleder 1996) of a previous generation of technology.

Despite the preplanning, changes were required to be made to the LIDET database schema during the experiment. The original database schema was planned according to the needs of 21 sites. When seven additional sites joined, the number of collected samples increased causing overlapping of sample identifying numbers. The solution was to change the 4-digit identifying number to a 5-digit number in the database schema. A member of the LIDET personnel points out the effort and contribution that was needed because of the tag number change in the database:

"... some of the tags had to be changed, and then the new tag numbers presented different problems with being able to merge the site information with the lab information ... it just makes for a very, very tedious effort to pull everything back together..." (LIDET personnel)

This change of the identifying number during the LIDET-experiment also required changes to the metadata descriptions. Another example was when an additional comment field for pooled samples and a flag field for outliers related to identifying a tag were added in the database. Additional codes also had to be made for some divergencies recognised in the samples. These kinds of issues during the long-term experiment required changes to the database schema. This demonstrates the importance of flexibility and tailorability of the information infrastructures in LIDET.

3.1.3. Metadata and auxiliary data related to research

Preserving the usability of scientific data in the long term requires data to be accompanied by contextual information that describes the data collections. These descriptions are called metadata (data about data). (Karasti and Baker 2004.) Information can be lost through degradation of the raw data or metadata. Adequate documentation (metadata) of sampling and analytical procedures, data anomalies, and data set structure will help to insure that data can be correctly interpreted or reinterpreted at a later day. (Michener et al., 1997.) Bowker and Star (1999) further suggest that formal data descriptions "wrapped" in informal descriptions increase the usefulness of the data.

Information concerning the LIDET-experiments stored in the FSDB includes the data and related metadata comprising an abstract describing the experiment, the formats and variable definitions of the data, and the programs used to process the data.

One example of metadata in LIDET was sample site information. Collaborators were asked to draw a map and write descriptions on the location of the research site and samples. When the permanent plots are set up it is essential to attach instructions on how to be able to find the samples. A member of the LIDET personnel also advised field collaborators to document unexpected samples:

"... there was a rock in here, or the bag was ripped, or animals seemed to have opened this bag " (LIDET personnel)

Relating this kind of auxiliary data in the database proved its value already during the experiment when researchers detected anomalies and assured the quality of measurements.

Yet another example of metadata is climate data. Weather is an important parameter in many site and synthesis studies. LTER climate database (CLIMDB) provides current and comparable climate summaries for each LTER site(Baker et al. 2000). Harvested climate data did not completely cover all LIDET sites and was not updated for all sites. Field collaborators of LIDET agreed to provide site-specific background information on climate, but getting them to provide other auxiliary data sets was more difficult to produce because of limited resources. A member of LIDET personnel explains his views about problems in gathering the related data:

" ... gathering of sort of this not extraneous but related data that wasn't really collected in LIDET and all that, it's been a problem. But I think it probably indicates a more classic effect of trying to work on the network when people can't agree on priorities, resources are limited..." (LIDET personnel)

Even though the information infrastructure in LIDET was designed to include metadata for long-term research use, the available resources were a limiting factor in the gathering and preserving of related contextual metadata.

3.1.4. Merging with other databases and information systems

Merging differently tagged pieces of information from two different databases of LIDET experiment also demanded the contribution of data management personnel. Tagged sample identification information was changed during the experiment, causing compatibility problems when the site-specific field data was merged with the laboratory measurements such as weight values.

Merging the LIDET-database with other LTER databases was the responsibility of the information manager. The documentation and data of the LIDET-experiment was planned to be stored along with LTER and other ecological data sets at FSDB (Stafford et al., 1988). The broader view and greater network level knowledge of the experienced information manager of LTER information systems were required in order to merge data of the LIDET-database with FSDB.

Concerning climate data, local technological infrastructure of LIDET differed from LTER network level infrastructure. Getting background climate data from the LTER climate database (CLIMDB) and merging this data with LIDET database required extra work from information management. Instead of being able to use the User Interface of CLIMDB the information manager harvested the climate data straight from CLIMDB and then transferred the data to the LIDET-database. According to Star and Ruhleder (1996) infrastructure takes on transparency by plugging into other infrastructures and tools in a standardized fashion. This points to the importance of compatibility in infrastructures when merging site and network level data.

3.2 Distributed experiment and collaboration

Information management of distributed and multi-intersite experiments like LIDET required contribution to both coordination and organisation work. Plans in LIDET were used more as resources for situated actions (Suchman 1987) in the sense that the centralized data management and organising of the distributed experiment allowed modifying the study plans in order to meet specific needs in the local situations. This in turn required effort and embedded knowledge from data management personnel.

By annual mailings and preparation of “study in a box”- research materials, the participation of geographically distributed field collaborators was made as effortless as possible. A member of the LIDET personnel describes the annual study preparations:

“… a lot of that organizational work was done … by us and it made it real easy the field cooperators to … study in a box … if we had sent them the loose bags and string and more detailed instructions (and told them to buy flags and …

that kind of thing) … I'm sure it would have resulted in more fallout, loss of cooperators. … " (LBC)

Managing a large number of samples collected from distributed sites was challenging. The massive number of sample boxes and samples to be handled demanded spatial organization and agreed procedures for the preservation and handling of samples in the laboratory. All the research material at a certain phase of analysis was placed in a certain place in the laboratory. This organizational effort also made the infrastructure visible (Star & Ruhleder 1996) to temporary staff, who could then learn correct work practices.

Previously communication infrastructure was based on personal telephone calls and letters. The technological development of Internet and e-mail made mass mailings possible and made communication with field collaborators less time-consuming. Also the sharing of information via www-pages became possible. This advantage to coordination of distributed experiment is highlighted in an interview:

"… in terms of coordination, this was being set up just as the internet, email , and all that was starting to be coordinated by the network office, so when I started this out I had to call all these people. But within a couple of months … I could do mass mailings. "(LIDET personnel)

LIDET infrastructures were built to support widely distributed scientific collaboration enabling the collaboration between a central site and geographically remote field sites. The wide geographical distribution of the participating sites defined the spatial scope property of infrastructure (Star & Ruhleder 1996).

3.3 Data sharing and publishing

Internet technologies offer opportunities for data sharing on an extensive scale (Karasti et al 2004). In LIDET guidelines for sharing data and publication were agreed by the LIDET participants. Data from the LIDET-experiment is already available on the Internet for intersite analysis and for model parametrization and testing.

Concerning the issues of data sharing and data reuse, Karasti et. al (2004) summarize how various examples show that long-time datasets exist but that they are frequently incomplete, badly maintained and not well documented. Data exist but is not useable by any but the local user, requiring too many resources for quality assurance and control (Karasti et al. 2004).In LIDET, test analyses of data were made to control the quality of collected data before final analysis. Test analyses, in addition to providing preliminary results of data, gave feedback of infrastructure operation and would have pointed out areas of improvement in information practices. In order to succeed in the long-term experiment a member of the LIDET personnel also highlighted the use of data before the final analysis:

"... You just have to use the data. Even if you're not doing the final analysis you have to sort of exorcise it to see in fact what it should be. Otherwise, you will end up with data, you won't why it's the way it is. And you won't be able to correct it 10 years later" (LIDET personnel)

Even though in terms of the publications the LIDET-experiment had a very elaborate arrangement, less data than could be expected has been published. A summarising article of the results of different LIDET teams has been published, but on the other hand many of the investigators have delayed the release of data. On reason presented by Helly et al. (2002) is that data submitters protect their data from being scooped by a quicker or alternative interpretation of their own data and consequent publication. A proposal of a mechanism to encourage and enable early publication of scientific data in a manner that produces beneficial side effects has been presented in an article on publication of digital scientific data by Helly et al., 2002.

4 Summary

A long-term, distributed ecological LIDET-experiment was initiated to enable the study of different ecosystems with standardized methods in geographically widely distributed sites. The long-term and distributed nature of the experiment caused challenges for the data management.

Through qualitative content analysis of the material it has been possible to study how information practices are intertwined with ecological research work practices. Everyday information practices of data management of the LIDET-case were studied applying a socio-technical approach (Kling 1999). The notions of infrastructure (Star & Ruhleder 1996, Star & Bowker 2002) allowed to analyse the information practices and collaboration infrastructure of the case in a socially sensitive manner. This paper foregrounds the long-term perspective and the elements of data management and information practices of LIDET-experiment.

Through the analysis of the empirical material of the LIDET case there is an emphasis on the importance of data management preplanning and the intertwining of research and data management practices. Data management design of the LIDET-experiment was driven by the research. The most important features of the LIDET-database were defined and maintained in collaboration with experienced ecologists, information managers and information systems developers.

Based on the interview material it has been possible to notify and capture some of the embedded knowledge of LIDET participants entrenched in the research and data management practices of the described ecological experiment. This knowledge that members of the CoP had was used in the design and modifications of the information infrastructure in the LIDET-experiment.

Management of research procedures as well as centralized handling of research samples and data in one laboratory have been the basic solutions to the data management challenges. The centralized preplanning, organisation and coordination of data management minimized the collaboration effort of the participating research sites in order to maintain the social infrastructure of collaborators. Only a few sites disappeared during this long-term experiment and the losses of information during the experiment phases were also minor.

The use, maintenance and development of long-term collaboration infrastructure of LIDET was a gradual process including modifying and tailoring of infrastructure in order to support the research needs and assure the long-term preservation and reuse of data. Improvements to everyday working practices shaped the infrastructure, and changes in the number of collaborating sites created a need to change the LIDET database. Infrastructure changes were also brought by new technologies developed during the experiment. New technologies offered new possibilities, e.g. in communication.

The technological development of LIDET information infrastructure was also prioritised according to its support for long-term science and data archiving purpose. In some cases old technology co-existed with new technology until the end of experiment.

Specific preplanning of the experimental design and anticipation of possible problems during the long-term formed the basis for intertwined design of the data management and infrastructure. The information infrastructure of LIDET-experiment, created by collaborative preplanning and centralized data management, was able to flexibly adjust to changes and thus succesfully supported scientific collaboration in the LIDET-case.

Based on this research there is need to carry out more studies of the use, maintenance and development of scientific collaboration infrastructures for the long-term perspective through other case studies. Further research is needed to increase understanding of how to relate local information practices and network-level solutions to be implemented in common infrastructures.

Acknowledgements

I want to express my thanks to professor Helena Karasti for guidance. I acknowledge EnviroNetGrants, multidisciplinary doctoral level Graduate school in Environmental Issues at the University of Oulu for support this work.

References

Baker, K., Bowker C. & Karasti, H. (2002). Designing and Infrastructure for Heterogeneity in Ecosystem Data, Collaborators and Organizations. Procs of the 2nd National Conference on Digital Government Research, 20-22 May 2002, Los Angeles, CA., Digital Government Research center, pp. 141-144.

Bowker, G. & Star, S. (1999). Sorting Things Out: Classification and its Consequences MIT Press, London.

Bowker, G., Star, S., Turner, W., Gasser, L. (eds.). 1997. Social Science, Technical Systems and Cooperative Work: Beyond the Great Devide. Mahwah (NJ): Lawrence Erlbaum Associates.

Callahan, T. (1984). Long-term Ecological Research. Bioscience 34(6):363-367.

Catanzaro, M. (1988). Using qualitative analytical technigues. Pp. 437-456 in Woods N. - Catanzaro M. (eds.) Nursing Research: Theory and Practice. C.V. Mosby Company.

Finholt, T.A. (2002). Collaboratories. In Cronin, B (ed.). Annual Review of Information Science and Technology. Medford, Information Today. Annual Review of Information Science and Technology, 36, 73-107.

Franklin, J. , Bledsoe, C. & Callahan, J. (1990). Contributions of the Long-Term Ecological Research Program, An expanded network of scientists, sites and programs can provide comparative analyses.

Gosz, J. (1999). Ecology Challenged? Who? Why? Where is This Headed? Ecosystems (2), 1999, pp.475-481.

Helly, J., Elvins, T., Sutton, D., Martinez, D., Miller, S., Pickett, S. & Ellison, A. (2002). Controlled Publication of Digital Scientific Data. Communications of the ACM. May 2002, Vol. 45. No. 5

Hobbie, J. et al. (2003). Scientific Accomplishments of the Long Term Ecological Research Program: An Introduction. BioScience, Vol. 53 No.1.

Kanfer, A. et al. (2000). Modelling Distributed Knowledge Processes in Next Generation Multidisciplinary Alliances. Available: http://citeseer.nj.nec.com/339903.html.

Karasti Helena 2001: Bridging work practice and system design - integrating systemic analysis, appreciative intervention and practitioner participation, Computer Supported Cooperative Work – An International Journal, Vol. 10. No 2, pp. 211-246.

Karasti, H., Baker, K. & Bowker C. (2004). ECSCW 2003 Computer Supported Scientific Collaboration (CSSC) workshop report. SIGGROUP Bulletin.

Karasti, H. and Baker, K. (2004): Infrastructuring for the Long-Term: Ecological Information Management, Hawaii International Conference on System Sciences 2004 (HICSS'37). January 5-8 2004, Hawaii, USA.

Kareiva, P., & Anderson, M. (1988). Spatial aspects of species interactions: the wedding of models and experiments. Pp. 35-50 in A. Hastings, ed. Community ecology. Springer-Verlag, New York, New York. USA.

Kling, R. (1999). What is Social Informatics and Why Does it Matter? D-Lib Magazine 5(1). Available: www.dlib.org/dlib/january99/kling/01kling.html.

Krippendorff, K. (1980). Content analysis: An introduction to its methodology. Beverly Hills. Sage.

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

Levin, S. (1992). The problem of pattern and scale in ecology. Ecology 73:1943-1967

The Long-Term Intersite Decomposition Experiment Team. (1995). Meeting the challenge of long-term, broad-scale ecological experiments. LTER Publication No. 19. U.S. LTER Network Office, Seattle, WA.23 pp.

Michener W. (1986). Data management and long-term ecological research. Pp 1-8 in Michener W. , (ed.) Research Data Management in the Ecological Sciences. Columbia, S.C.: University of South Carolina Press.

Michener, W., Brunt, J., Helly, T., Kirchner, T. & Stafford, S. (1997). Non-geospatial metadata for the ecological sciences. Ecological Applications 7(1): 330-342.

Michener, W. K. & Brunt, J. W. (eds.) (2000). Ecological Data: Design, Management and Processing. Oxford, Blackwell Science. Methods in Ecology.

Miles, M. & Huberman, A. 1984. Qualitative data analysis: A sourcebook of new methods. Beverly Hills, Ca.: Sage.

Porter, J. (1998). Scientific databases for Environmental Research. Data and Information Management in the Ecological Sciences: A Resource Guide. Albuquerque, NM: LTER Network Office, University of New Mexico.

Porter J., Callahan, J. (1994). Circumventing a dilemma: historical approaches to data sharing in ecological research. Pp. 193-202 in Michener, WK, Brunt, JW & Stafford, SG, eds. Environmental Information Management and Analysis : Ecosystem to Global Scales.

Stafford, S. (1998). Issues and consepts of Data Management: The H.J. Andrews Forest Science Data Bank as a Case Study. Data and Information Management in the Ecological Sciences: A Resource Guide. Albuquerque, NM: LTER Network Office, University of New Mexico. Available: www.ecoinformatics.org/pubs/guide/stafford.fv2.htm

Stafford, S., Spycher, G. & Klopsch, M. (1988). Evolution of the forest science Data bank. A progress report on managing complex research data. Journal of Forestry, Vol. 86, No.9, September 1988.

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

Star, S. & Bowker, G. (2002). How to Infrastructure. In Lievrouw, L. A. & Livingstone, S. (eds.). Handbook of New Media : Social Shaping and Consequences of ICTs. London, Sage.

Strauss, A. and Corbin, J. (1990). Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Newbury Park, CA: Sage Publications.

Suchman, L. (1987). Plans and situated actions: the problems of human-machine communication. Cambridge, Cambridge University Press.

Tuomi, J. & Sarajärvi, A. (2002). Laadullinen tutkimus ja sisältöanalyysi. Tammi, Helsinki.

Woolgar, S. (2003) Social Shaping Perspectives on e-Science and e-Social Science: The case for research support. A consultative study for the Economic and Social Research Council (ESRC). (Downloaded 27 June 2004 at http://www.sbs.ox.ac.uk/downloads/E-SocialScience.pdf).