THE LTER NETWORK INFORMATION SYSTEM:

A FRAMEWORK FOR ECOLOGICAL INFORMATION MANAGEMENT

James W. Brunt

Access to data and metadata distributed throughout collaborating national and international research sites is the focus of a development effort to support ecological research. The U.S. LTER (Long Term Ecological Research) Network informatics group is developing, along with national and international partners, a Network Information System (NIS). The mission of the working group is to design and develop a modular framework that builds on existing site functionality. Platform independence and the capability to present data and information in a consistent fashion, independent of their original format and location, are essential elements of design for the system to function in heterogeneous site environments. The key element, however, is the use of interdisciplinary NIS task groups composed of scientists from existing research groups, information managers, and computer scientists to design and implement various components of the system. Current prototypes serve as data sharing models for improving access to ecological data across sites.

INTRODUCTION

Design and implementation of an information system that seamlessly facilitates intersite research is a major challenge to the U.S. LTER network in the coming decade. The necessity of intersite research has driven the LTER mandate since workshops in the 1980's (Marzolf and Dyer 1986). Since its organisation in 1988, the LTER information managers group has focused on facilitating this research. Although computer technology and ecological informatics have come a long way, the fundamentals of managing research information and developing an appropriate system for management have remained relatively stable. What have changed drastically in the last decade are researcher and community expectations. With the explosion of Internet connectivity and the birth of the World-Wide Web (WWW), scientists, administrators, and the general public have come to expect greater access to the products of valuable research dollars (Stafford, et al. 1994). The LTER research network has a wealth of long-term data that are being collected by LTER investigators. With the publication of the LTER Catalog of Core Data Sets in 1990, this resource became more widely known within the ecological research community.

THE LTER NETWORK INFORMATION SYSTEM (1990-1995)

During the same period that the LTER information managers developed the catalog of core data sets, the LTER Connectivity Committee established nearly full Internet connectivity across the U.S. LTER Network. This infrastructure for electronic communication and data exchange substantially enhanced the traditional means of networking. Based on this rapidly evolving infrastructure, electronic versions of the Core Data set Catalog and the LTER Personnel Directory were made available to most researchers in 1991(Table 1). The evolution of those data repositories from hardcopy versions to on-line databases accessible by e-mail, FTP, Gopher, WAIS, SQL and Web information servers reflects the successful use of emerging communications technology in the integration of individual sites into a coherent research network. When the LTER All-Site Bibliography was created in 1993, a hardcopy version was no longer considered an economical or even desirable implementation (Chinn and Bledsoe, 1997). Instead, it was implemented as a searchable on-line database.

Table 1 - History of LTER Network Information System development 1988-1998

1988	Minimum Standard Installation (MSI)
1988	BITNET email
1989	LTERnet Internet-based e-mail forwarding system
1990	Internet Connectivity

1991	Gopher servers
1992	WAIS servers
1993	HTTP servers
1994	On-line Data
1995	Web to Database Connectivity
1996	NIS plan draft
1997	Prototypes developed
1998	Prototypes being evaluated

The improved accessibility to data and ease of use of web browsers highlighted gaps in the LTER information infrastructure. Gopher and Web servers simplified access to widely dispersed LTER information, but queries were still unwieldy. The All-Site Bibliography, for example, although on-line and seeing increasing use eweeee has limitations in the type and number of queries that are possible. Most LTER data sets are available via WWW and other network information servers, but in many cases to find out what data are available on LTER site information servers, you need to make connections to each of the site servers individually, and it is difficult to analyze and synthesize data from different sites because most site servers return data in different formats. The LTER information managers have formulated their vision for an integrated network information system (NIS) that will alleviate these and other difficulties with LTER data access. An NIS working group has been established to guide and advance the process to develop and to implement the vision.

PHILOSOPHY AND GOALS OF LTER NIS (1996-2002)

The fundamental philosophy of the effort is "**Develop from a Research Perspective.**" The primary function of the NIS will be the support of intersite research. Therefore, development of the NIS is closely linked to efforts by LTER cross-site and synthesis research groups (Figure 1). Each intersite research effort includes at least one member from the LTER information management group. Initially, intersite research groups provide the main pool of users of new system functions and they provide the feedback required for testing. In addition, development is made with the recognition that information management systems are always evolving, due both to technological changes and to improvements in our understanding of how scientists most efficiently use data. The LTER NIS effort is integrating site information systems, not replacing them. The envisioned system will continue to take advantage of advances in network tools that ensures use into the coming decade.

The goals for this development effort are:

Goal 1: Increase the utility of the existing system

Some basic components of a network information system are available to LTER researchers (eg. catalog, personnel directory, bibliography, integrated mail forwarding, and direct links to most site systems). A modular, step-by-step approach to the NIS development ensures that existing functions will be seamlessly integrated with the future system. With the objective of fulfilling its mission in information management, the LTER Network Office will be responsible for further development of the LTERnet Network Support System to maintain and expand the following core activities:

- LTERnet WWW server
- Personnel Directory
- Core Dataset Catalog
- Cross-site Bibliography
- Satellite Imagery Archive

- Electronic Connectivity
- Query and retrieval systems to integrate these functions

Goal 2: Increase access and query capability on intersite data

The Network Information System working group has developed the following query and retrieval functions for the NIS based on input from the data management committee and researchers. An advanced access and query capability will facilitate:

- a search for data available anywhere in the LTER network,
- combining and analyzing data from different sites,
- answering standard information requests (those requests that occur >80 of the time)
- economically building query systems for specific projects and special information requests
- automatic retrieval of data and documentation
- query, analysis and display tools that are intuitive to researchers.

Goal 3: Maintain strength in diversity

Research Data Management at U.S. LTER sites is carried out with common objectives in mind, but under a variety of different circumstances and constraints. There are a number of heterogeneous approaches taken to meet research objectives (Conley and Brunt 1991, Briggs and Su 1994, Brunt 1994, Ingersol et al. 1997). Meeting standardized goals with a variety of solutions has built strength into this system. This strength is being maintained and nurtured in the development of the NIS.

NIS DEVELOPMENT STRATEGIES

Design and implementation of the NIS is being done in a modular, step-by-step fashion, using prototypes to evaluate different technologies, and involving subsequent evaluation for interoperability. These strategies require that teams of information specialists accomplish specific system parts in a coordinated way and in a predictable period. Such tasks include implementation of access and retrieval functions for data sets that are standardized in terms of content and apparent format. For example, measurement procedures of net primary productivity can vary considerably among such different ecosystems as lakes, forests, deserts, and grasslands. Correspondingly, the raw data reflecting those measurements may not be comparable. In these cases, a team of researchers and information specialists will work with the sites to add to the NIS the capability to access from all sites productivity data and metadata in a consistent format, with consistent tools.

COORDINATION

The development of an LTER-wide Information system depends on the efforts of many critical players: The LTER Coordinating Committee, site researchers, Network Office Staff, Data/Information Management Committee, and Collaborators (e.g., SDSC). These efforts are focused at a continuing series of workshops that form the basis for the design, funding, and implementation of the LTER NIS. Annual face-to-face meetings of the Data Management Committee and the NIS working group along with frequent conference calls serve to coordinate the effort. NIS task groups composed of researchers, data managers, and programmers provide the insight and labor necessary to design, fund, and implement various components of the system.

TASK WORKSHOPS

The critical element in the NIS development is the use of interdisciplinary task groups composed of scientists from existing cross-site research groups, information managers, and computer scientists to design and implement specific components of the system (Figure 1). The NIS working group solicits interest from intersite research groups and holds workshops to assess researchers needs. These workshops have been organized in conjunction with workshops that investigate specific intersite research questions, such as climate patterns, biodiversity distribution, litter decomposition processes or soil sampling methods (see: Conley 1991). The LTER Coordinating Committee determines which areas to cover first. The workshops address and attempt to resolve outstanding standardization issues that are relevant to how

data and information are presented and distributed by the NIS. Data of interest to the group, such as the Climate Committee have been used by a task group to demonstrate present capabilities. For example, previously, to compile a web page of all climate data available on LTER site servers, one would download them from individual site servers into spreadsheets or other applications, and graph them. The task group evaluates these brute force approaches in use for what works and what doesn't work. Why are these methods insufficient in supporting the groups intersite research? The group then proposes solutions to fix the shortcomings, and sets a timetable for implementation of a prototype. Prototype development includes developing an exchange format, information model for the data, and a WWW interface to the data. The process would then proceed with testing and feedback. The positive aspects of the task group experiences are then used in other working groups.
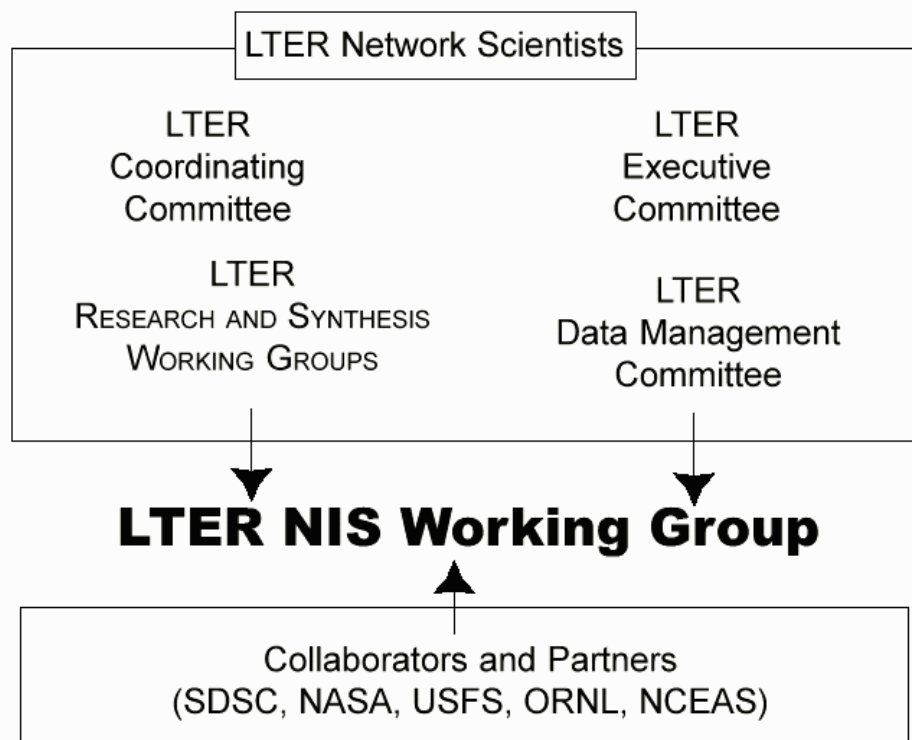

Figure 1 - Diagrammatic representation of the interaction of groups involved in the LTER NIS development.

STANDARDS DEVELOPMENT AND USE

The key to developing a truly interoperable information system is the use of a flexible metadata model that is rooted in scientifically meaningful content. Standards such as the U.S. Federal Spatial Data Transfer Standard (SDTS), designed to facilitate communication of accurate information, fail to meet the expectations of the "20-year rule" for reuse (Figure 2). This criterion, imposed by LTER researchers for preservation of complete information, states that data need to be documented such that independent scientists can understand and use the data 20 years after collection (Porter and Callahan 1994). This is necessary because of the degradation of information that occurs with incomplete documentation described by Michener and colleagues (1997). LTER information managers are continually involved in scrutiny of new standards and software being adopted elsewhere for possible implementation in the LTER Network (e.g. FGDC (FGDC 1994), Dublin Core (Weibel 1997), etc.). They have also been active in the recent formalization of content standards for non-spatial ecological data (Michener, et al., 1997).
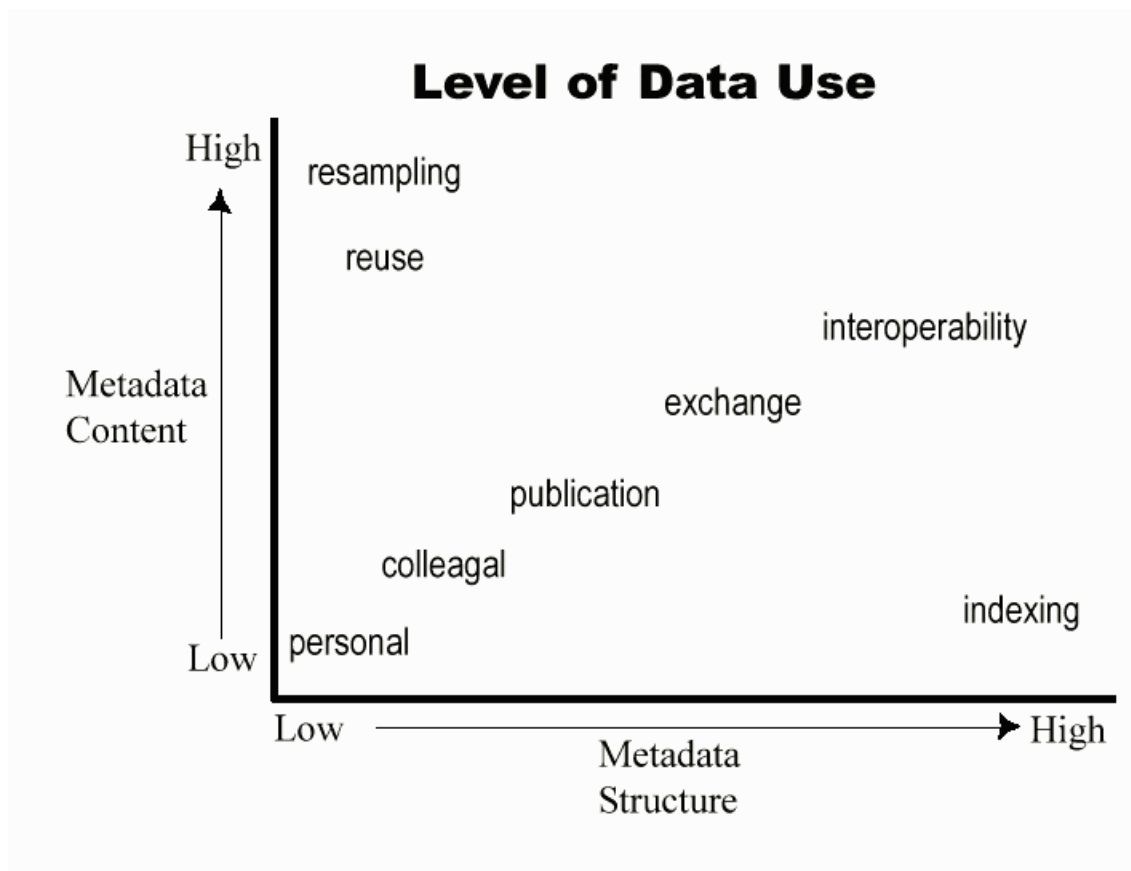
Level of Data Use

Figure 2 - The information content of metadata necessary to re-use data 20 years after collection is higher than that required for exchange of data. This demonstrates the need for broader, more flexible metadata for ecological data than is provided for in current standards.

Industry and de facto standards for network data communication that can help ensure hardware and software independent functionality have emerged and are being considered in the NIS design. For example, network client/server technology using ODBC is well suited to this application and can be implemented independent of the specific computer platform. There are also some recent developments with potentially major implications for the way future network information systems are implemented. In particular, executable content on Web servers by the Java language (an open standard for programming on the Web) could be used to package data together with their metadata - a requirement introduced by Conley and Brunt (1991).

NIS PROTOTYPES

The NIS prototype developments take advantage of advances in web-based database interfaces to provide access to network-wide data sets via a single point of entry. A number of prototypes have been developed addressing different data types found in the network. Each of these prototypes was developed using a different database/web technology for comparison. The prototypes include the climate databases (project name: CLIMDB), the site description information (SITEDB), and a basic data catalog (DTOC), and the personnel database (PERSDB). The prototypes are being developed and tested at LTER sites prior to being installed on the new LTER Network information system infrastructure (LTERnet) at the Network Office.

CLIMDB

The climate database (CLIMDB) prototype is an interactive, web-based, interface to standardized LTER Network climate data (Henshaw et al. 1998). To develop CLIMDB, LTER information managers engaged with LTER climate researchers, in a relationship where current database technology was applied to the climate data standards (see: http://www.geog.unc.edu/~greenland/CLIMSTAN.html/Default.htm). The methods used to bring these two groups productively together are now being refined in such a way that they can be applied to other, more ecologically oriented data sets that will eventually become components of the LTER NIS. Plans are now in progress to migrate CLIMDB,

from an ORACLE database supported by the North Temperate Lakes LTER in Madison (http://www.limnology.wisc.edu/climdb.html , to Microsoft SQL Server at the LTER Network Office.

DTOC

The Data Table of Contents (DTOC) is an automated, web-based data catalog for LTER site data sets. It contains a streamlined, standardized metadata set related to each research data set. This information is "harvested" from the site information systems, using similar conceptual methods to those developed in the CLIMDB project (Figure 3). This project provides the groundwork for the implementation of a cross-site, standardized metadata model that would be the backbone of the NIS Interoperability Framework. The Data Management Committee, with the support of the LTER Network Office, developed this project which was housed on the Virginia Coast Reserve LTER site information system. The DTOC has just been moved and is being tested on the LTER Network Office Computer system (http://www.lternet.edu/DTOC/).
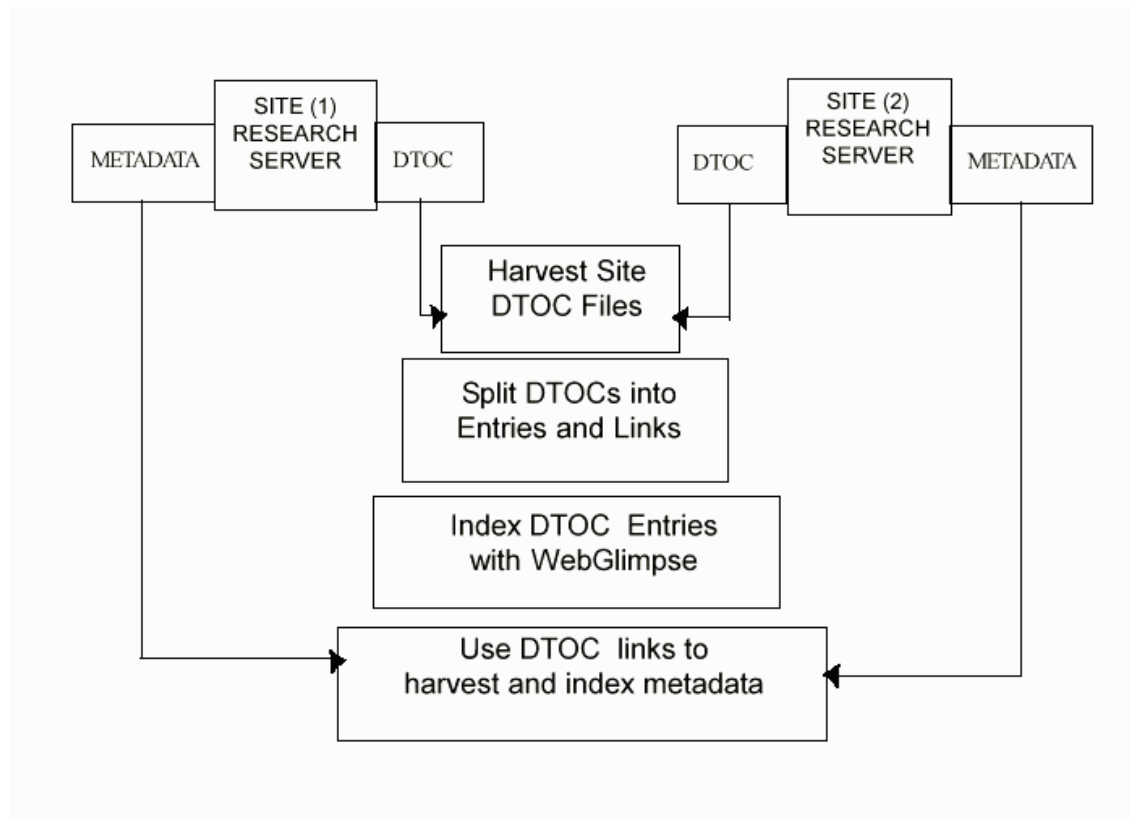

Figure 3 - The flow of information in the preparation of data table of contents (DTOC) indices. The harvesting server retrieves the data set entries along with URLs to more extensive metadata and data which can be accessed via the WWW searchable index using WebGlimpse

SITEDB

The site description database (SITEDB) project adds another level of metadata compliance to the NIS. This work has been prompted both from a scientific and an administrative need for detailed, descriptive information about each LTER site. SITEDB presents standardized web page views for each site in addition to having all the relevant site information available in a single relational table. This information can be maintained by the site via web-based interface. The implementation is in miniSQL at the Palmer Island LTER headquarters in San Diego and it too will eventually be integrated into the NIS.

PERSDB

The LTER Network personnel database has been revised and migrated to MS SQL from an ingres database. Among the improvements is the ability to more closely control entries and mail lists, and integration with the developing LTER NIS. This database can now be managed updated by the sites via a web-based interface implemented in PERL.

All these advances were made both through financial support of the Network Office for work of the LTER Data Management and Climate Committees and by direct participation of NET and site personnel.

NIS INTEROPERABILITY FRAMEWORK

The group is currently in the process of evaluating the prototype modules, described above. The results will be developed into the LTER NIS interoperability framework – a set of specifications that describe the interoperability of the discrete units of the system, into which additional, future modules can be plugged (Figure 4).

COMPONENTS OF THE FRAMEWORK

- Distribution Format Specification
- "Centributed" Harvester Specification
- Indexer Specification
- Metadatabase Specification
- Server Specification
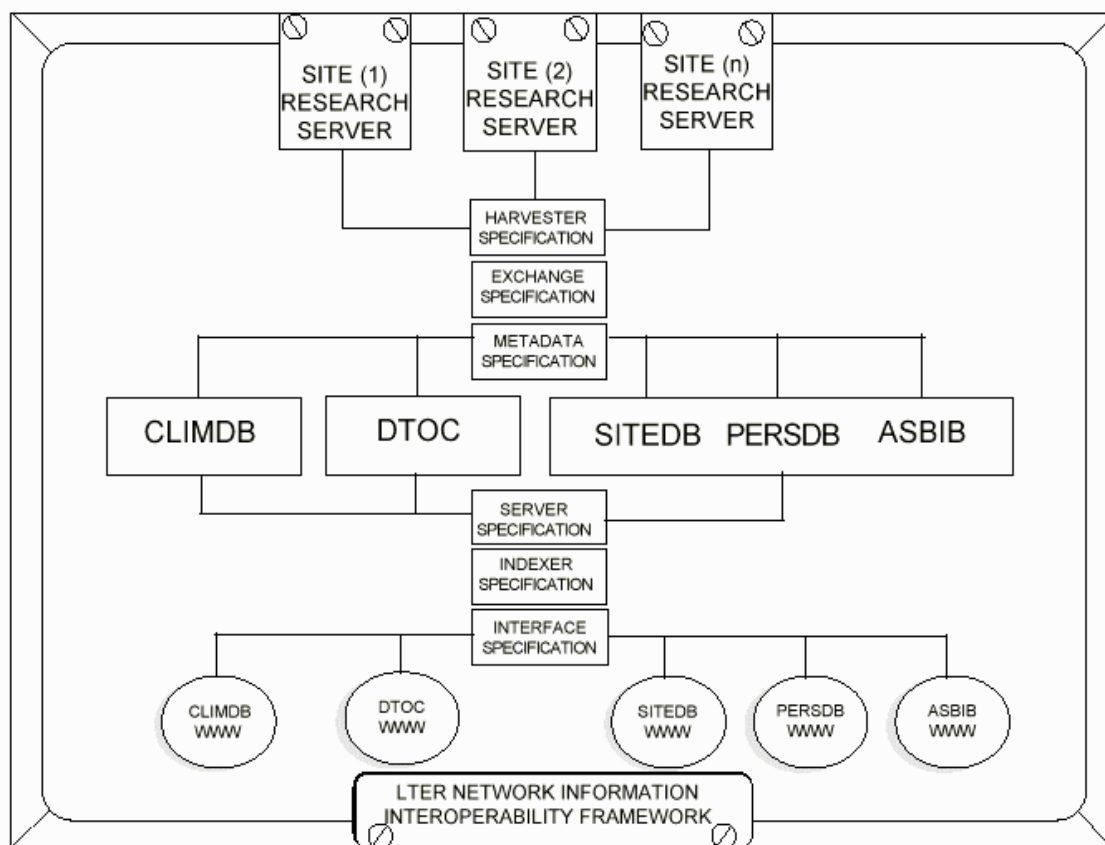- WWW Interface Specification

Figure 4 - Diagrammatic representation of the relationships of the specification components of the LTER Network Interoperability Framework. The Framework provides the structure for specifications connecting LTER site information modules with the LTER Network Information System. As envisioned, the system is completely modular allowing for databases and sites, to be plugged-in at any time.

RESOURCES

The development of a distributed NIS depends upon the readiness of individual sites to participate but will not exclude the site's information. Sites that do not yet have full capabilities can have their contributions supported by the Network Office server. The LTER Network Office is charged with supporting the expansion of the existing NIS and advanced query and information systems that integrate data from the individual site information systems. The Network Office also supports planning activities, aid in coordination of site activities, promote standards development, develop

Network data sets, and provide access to software, storage and network resources.

NETWORK INFORMATION SYSTEM INFRASTRUCTURE

The backbone of the NIS is the computational infrastructure at the LTER Network Office. Recently installed scalable servers and increased network bandwidth will better serve the LTER Network and ecological community. Two Sun 450 Enterprise servers and two Dell Poweredge servers are the primary computers serving LTERNET. The combination of Sun Solaris operating system on the UltraSparc platform and the Windows NT operating system on the Intel platform allows for maximum flexibility in incorporating new developments and technology, each with strengths that compliment the other. The LTER Network Office is the process of upgrading its connection to the vBNS internet to Gigabit speeds (1000 Mb/s).

INTERNATIONAL INFORMATION MANAGEMENT INTEGRATION

LTER information managers are actively cooperating with several international efforts to assure interoperability between the LTER Network and the greater scientific community. LTER Network Office personnel and Network Information managers participated in a data management working group as part of the ILTER meeting June, 1997 in Brazil. This workshop led to the development of a planning meeting that was held in early December in Albuquerque hosted by the LTER Network Office to prepare Latin American Information Management experts for a data management training effort in Venezuela in June, 1998. The June meeting brought together data managers from LTER sites all over Latin America and established the groundwork for developing international standards for data exchange and interoperability and will provided exchange of information that will be valuable to developing ILTER networks.

FUTURE DEVELOPMENTS

Activities for the future will focus on the development of the Network Information System Interoperability Framework for connecting LTER site information modules with the LTER Network Information System. Organizational activities have begun towards developing the Nitrogen Deposition module for LTER NIS and discussions are in progress on other LTER Network data areas for integration. All these efforts include the implementation of broad metadata interoperability specifications.

LITERATURE CITED

Briggs, J. & Su, H. 1994. Development and refinement of the Konza Prairie LTER research information management program. In: *Environmental information management and analysis: ecosystem to global scales* (eds W. K. Michener, J. W. Brunt, & S. G. Stafford), pp. 87-100. Taylor and Francis, Ltd., London, England.

Brunt, J. W., & Brigham, W. 1992. Data standards for collaboration in science. In: *Data management at biological field stations and coastal marine labs: report of an invitational workshop* (ed J. Gorentz), pp. 15-17. Kellogg Biological Station, Kalamazoo, Michigan, USA.

Brunt, J. W. 1994. Research data management in ecology: a practical approach for long term projects. In: *Seventh international working conference on scientific and statistical database management* (eds J.C. French & H. Hinterberger), pp. 272-275. IEEE Computer Society Press, Washington, DC., USA

Chinn, H. and C. Bledsoe. 1997. Internet Access to Ecological Information: The U.S. LTER All-Site Bibliography Project. *BioScience* 47(1).

Conley, W. 1991. An Institute for theoretical ecology? – part IV: ``computational workshops'': a planned activity for theoretical ecology. *Coenoses* **6,** (2), 113-120.

Conley, W., & Brunt, J. W. 1991. An institute for theoretical ecology? Part V: practical data management for cross-site analysis and synthesis of ecological information. *Coenoses* **6**, (3), 173-180.

Federal Geographic Data Committee. 1994. Content Standards for Digital Geospatial Metadata (June 8). Federal Geographic Data Committee, Washington, DC, USA.

Henshaw, D. L., et al. (1998) Climate database project: a strategy for improving information access across research sites. In*: Data and information management in the ecological sciences: a resource guide* (eds W. K. Michener, J. H. Porter, & S. G. Stafford), pp. 29-31. LTER Network Office, University of New Mexico, Albuquerque, New Mexico, USA.

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

Marzolf, G. R., & Dyer, M. I. 1986. Future directions for research data management in ecology. In: *Research data management in the ecological sciences* (ed W.K. Michener), pp. 411-420 University of South Carolina Press, Columbia, South Carolina, USA.

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

Porter, J. H., & Callahan, J. T. 1994. Circumventing a dilemma: historical approaches to data sharing in ecological research. In: *Environmental information management and analysis: ecosystem to global scales* (eds W. K. Michener, J. W. Brunt, & S. G. Stafford), pp. 194-202. Taylor and Francis, Ltd., London, England.

Stafford, S. G., Brunt, J. W., & Michener, W. K. 1994. Integration of scientific information management and environmental research. In: *Environmental information management and analysis: ecosystem to global scales* (eds W. K. Michener, J. W. Brunt, & S. G. Stafford), pp. 3-19. Taylor and Francis, Ltd., London, England.

Weibel, S. 1997. The Dublin Core: a simple content description for electronic resources. *Bulletin of the American Society for Information Sciences*. October/November, 9-11.