LTER Information Managers 1998 Annual Meeting Report

i. Table of Contents

1.0 Executive Summary (S. Stafford (AND) and D. Henshaw (AND))

The annual LTER Information Managers Meeting was held July 30-August 1, 1998, hosted by the new, urban Baltimore LTER Site. Twenty of the 21 sites were represented with Cedar Creek not represented. Additional guests included Rick Clutter, National Estuarine Research Reserve System (NERRS), Baruch Institute; Cheryl Solomon, NASA Global Change Master Directory (GCMD) project; Dick Olson, Oak Ridge National Laboratory, DAAC project; and Miguel Acevedo, IM Committee for the Latin America Network.

Status reports were presented on the LTER Network Information System (NIS), including an NIS overview, the Data Table of Contents project (DTOC), the Climate Database project (ClimDB), and the Site Description Database project (SiteDB). The Network Office presented a detailed report on network activities, databases, personnel, International LTER (ILTER), KDI proposals, remote sensing activities, and the Global Terrestrial Observing System (GTOS). Other reports included knowledge management (or mechanisms on how we might organize ourselves to preserve our collective wisdom), the eXtensible Markup Language (XML), web publishing issues, and ESA publishing of peer-reviewed data papers. Special visitor reports included the Oak Ridge DAAC and the Latin American ILTER program.

Working groups focused attention on 1) future development of the NIS, 2) writing an explicit statement on the ethical use of online data, and 3) reviewing LTER Information Management (IM) outreach, Minimum Standard Installation (MSI), LTER 2000 meeting activities, and the Y2K problem. A final working session discussed the IM Committee's collaboration with the Northwest Alliance of Computational Science and Engineering (NACSE), and the writing of four pilot proposals to NACSE.

A closed LTER Data Manager session was conducted to review visitor participation at the annual LTER IM meeting and to conduct DataTask elections. The IM Committee also unanimously adopted the creation of a chairperson position to bring the LTER IM Committee structure in line with the structure of other LTER committees. Susan Stafford was unanimously elected as chair.

2.0 Network Information System Status (J. Porter (VCR))

The LTER Network needs to meet the ecological information challenge of providing information to ecologists in ways that let them locate the information they need, in forms they can readily use, and with the assurance that the information is current and accurate. To meet this challenge the LTER information managers in 1994 proposed developing a Network Information System (NIS, then called a Data and Information System [DIS]) (http://www.lternet.edu/documents/Reports/Data-management-committee/1994-DM-committee-report/im_1994_report.pdf). Discussed at the 1995 meeting was "The Dream Information System" which was described as follows: "*While sitting by pool, we ask our system 'Why is there biodiversity?' The system queries all site data servers, answers our question, and automatically submits a paper to Nature, which is immediately accepted..... then we woke up.....*" Although it is unlikely that all characteristics of the "dream" system will ever be realized, it nonetheless captures the essential purpose of the NIS: to promote ecological science by fostering the synergy of information systems and scientific research.

At the 1995 and 1996 meetings the objectives and missions of the NIS were elaborated:

- "The main objective of the present and future LTER NIS is to support basic ecological research and science, at both the site and network levels"
- "The mission of the NIS working group is to design and develop a distributed, LTER-wide information system using a modular approach, while maintaining and building on present functionality" (http://www.lternet.edu/documents/Reports/Data-management-committee/1996-DM-committee-report/im_1996_report.htm).

Since those meetings there has been significant progress in meeting NIS objectives using the modular approach and

extensive use of prototypes discussed in the 1995 report (http://www.lternet.edu/documents/Reports/Data-management-committee/1995-DM-committee-report/im_1995_report.htm). The current NIS components are shown in Table 1.

TABLE 1. Current NIS Components

Researcher Support Databases	**Description**	**Status**
• Personnel Database	Provides telephone, address and email information on all LTER researchers	Operational
• Topic-based Electronic Mailing Lists	Supports electronic mailing lists for LTER researchers	Operational
• All-site Bibliographic database	Free-text searches of LTER site bibliographies	Under Revision
Data-oriented Databases	**Description**	**Status**
• Data Catalog	Free text and keyword access to LTER site datasets	Operational
• All-Site Climate Database	Consistent interface to LTER meteorological data, with data selection and graphical capabilities	Operational Prototype
• Site Description Database	Database of basic site characteristics needed by most researchers	Prototype
• Nitrogen Deposition Database	All-site database of nitrogen deposition data	Planning

The framework of the NIS incorporates operational elements (located at the LTER Network Office) along with prototypes, which are frequently hosted by individual sites (Figure 1). For example, primary development of the Climate Database has taken place at the NTL LTER site (with standards development assistance from the AND site), but eventually the database will be ported to the LTER Network Office.

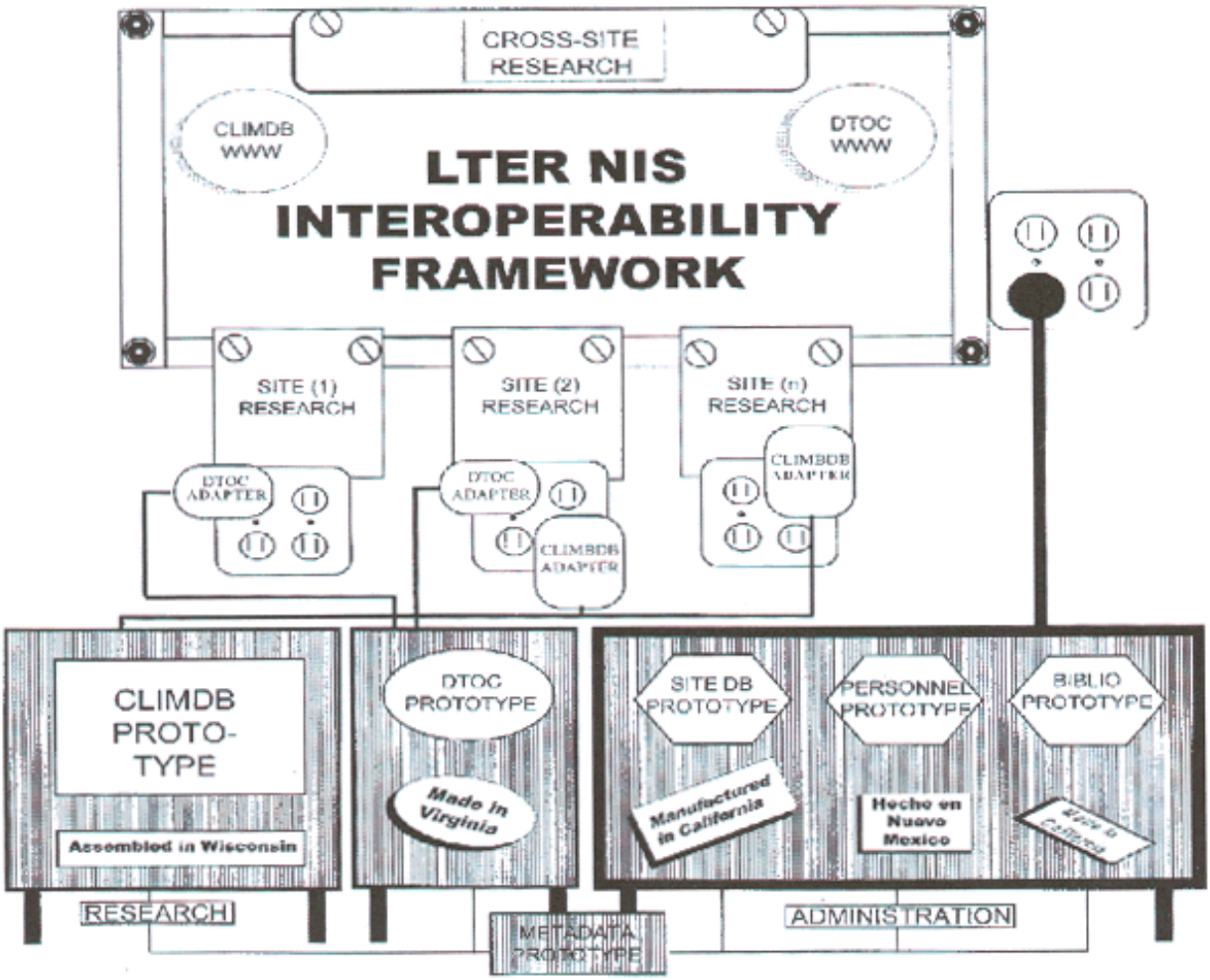

Figure 1. NIS Framework (from LTER Network News, 1998)

The framework takes full advantage of the diverse skills and capabilities at LTER sites by using the LTER network as a distributed information management laboratory. With 21 sites in the network, there are a large number of different approaches and types of software that can be simultaneously evaluated. Software and approaches that are successful can be "exported" to other sites. For example, the Climate Database drew on different expertise from different sites. Don Henshaw of the Andrews LTER used his expertise with meteorological data and connections with the LTER Climate Committee to take the lead on developing a draft exchange format for LTER climate data. Darrell Blodgett of the Bonanza Creek LTER site used his expertise in harvesting information across the WWW to help with the development of a protocol for making copies of data from the sites at a central location. Robin Stubbs and Barbara Benson then used their expertise with relational databases to create a system that supported ingestion, query and display using a WWW interface. In this instance, no individual site had all the expertise to build the system, but by combining the right elements from different sites, a highly functional system was created.

Components of the NIS are coming into use by a wider audience. Several existing modules of the NIS (personnel, mail-forwarding) are being used for the Global Terrestrial Observing System (GTOS) and the International Long-Term Ecological Research Network (ILTER).

2.1 Data Table of Contents Project Update

1997-1998 has been a period of consolidation for the Data Table of Contents (DTOC) Project. Currently 15 sites (AND, ARC, BNZ, CDR, CWT, HBR, JRN, KNZ,

LUQ, MCM, NTL, NWT, PAL, SGS, and VCR) participate in the data catalog. The prototype DTOC was described in the 1997 report (http://www.lternet.edu/documents/Reports/Data-management-committee/1997-DM-committee-report/im_1997_report.htm). The system provides a searchable, cross-site data catalog (http://www.lternet.edu/DTOC/) that is automatically updated on a weekly basis.

Recent tasks have focused on moving from a prototype to a fully operational system and extending functionality through the addition of "help" and keyword systems. Specific tasks were:

- Revising the programming code in the prototype DTOC to increase portability
- Moving the prototype DTOC to the LTERNET computer system http://www.lternet.edu/DTOC/
- Adding a "help" page to explain how to conduct complex Boolean searches
- Development of a keyword directory with links to individual datasets
- Adding additional sites to the DTOC system

At the 1997 meeting we discussed improving the DTOC by allowing complex Boolean searches and use of alternative search engines. Michael Hartman (NWT) investigated alternative search engines, but found nothing currently available that was significantly better than the WebGlimpse search engine currently in use. He discovered that WebGlimpse already incorporated Boolean searching, and wrote a "help" file describing the needed syntax. John Porter (VCR) did the code revision, porting to LTERNET and development of the keyword listing, with assists from James Brunt (NET) and other system operators at the LTER Network Office.

Several tasks were identified for the coming years. The highest priority is to add additional sites to achieve full network coverage in the data catalog. A second task is to refine the user interface and improve the number and quality of links to the DTOC by individual sites. The idea would be to provide a standardized "box" that could be incorporated into site WWW pages to provide search capabilities for both site and network datasets. The keyword system provides the first comprehensive listing of keywords in use at LTER sites, as well as providing links back to associated datasets.

However, it currently employs an *ad hoc* set of keywords provided by the sites, would benefit from the creation of a keyword thesaurus or semantic engine, which would link specific site keywords to a set of standardized archetypal keywords. This would allow a search for the word "rot" to return links to datasets on "decomposition." A final task is to work on linking the LTER DTOC with other data directories, such as the NASA Global Change Master Directory.

2.2 SiteDB: Prototype Site Description Database for LTER Sites (D.Blodgett (BNZ) and K.Baker (PAL))

Rationale:

As the LTER program matures, retrieval of basic site information across the whole network of sites becomes important. Recognition of the need for rapid access to general site description and history in addition to specifics such as latitude, longitude, area, elevation, and contacts stimulated interest in development of some type of site information database. Consideration is required with respect to integration with the LTER Network Information System (NIS) and with other site information systems such as the Terrestrial Ecological Monitoring Stations (TEMS) pilot project of Global Terrestrial Observing System (GTOS).

SiteDB Working Group :

Darrell Blodgett (BNZ)
Karen Baker (PAL)
John Anderson (JRN)
Dick Olson (NASA)
Chris Wasser (SGS)
Peter McCartney (CAP)
Ned Gardiner (CWT)
John Vande Castle (NET)

mailing list: sitedb@lternet.edu

SiteDB Action Items:

1. meeting report by Sept. 18 (Darrell, Karen)
2. distribute URL for SITEDB (Karen)
3. working group
    a. create working group with mailing list
       Darrell Blodgett (BNZ) (co-lead)
       Karen Baker (PAL) (co-lead)
       John Anderson (JRN)
       Dick Olson (NASA)
       Chris Wasser (SGS)
       Peter McCartney (CAP)
       Ned Gardiner (CWT)
       John Vande Castle (NET)
    b. Question1: review statement of purpose and goals (rationale)
    c. Question2: relationship with TEMS/GTOS
    d. Question3: create guidelines for contents
4. report of group by Sept. 30
5. individual data managers present SITEDB to site investigators
6. demo siteDB for EXEC and possibly CC feedback critical (JohnP)
7. invite further site participation

Prototype Initiated:
During the CLIMSTAN meeting at Sevilleta in October 1997, Karen Baker (PAL), Darrell Blodgett (BNZ), Barbara Benson (NTL), and Don Henshaw (AND) proposed a prototype for a Site description DataBase (SiteDB) that could be used quickly to find standard basic information about all LTER sites. A survey presented at the LTER Central Committee Meeting in October 1997 by Mark Harmon (AND) highlighted the difficulties of the existing system which depends upon browsing separately the LTER network server as well as all of the more than twenty-two individual site servers. An initial long list of parameters was shortened eliminating information that was either too detailed or too site specific. Keeping the content broad and basic maintains flexibility and addresses the concern that the time needed for participation be kept to a minimum in order to attract as much site participation as possible.

Methods and Modules:
A modular design provides an extensible organization for SiteDB. An initial prototype was developed using Mini-SQL (mSQL, Hughes) patterned after earlier successful Bonanza Creek projects implemented using this relational database software package. A hyper text mark-up language (HTML) file is run through the mSQL interpreter to translate mSQL commands into presentation views. Although all the information currently is stored in a single table, related information is used to create views or modules of particular interest. The main view, or site view, is the root module of the project. Further views, or modules, are topic specific such as climate, vegetation, regionalization, and soils. An early extension of the project was the addition of uniform resource locators (URL). Having important URLs in the database provides a single location for recording and maintaining links to additional individual site information on a variety of network projects including data in the form of harvestable files for the LTER climate database project (ClimDB), and possibly eventually for modules of SiteDB itself. Thus SiteDB not only provides an initial site

overview and themed views but pointers to further information as well.

Data Entry:
A short set of web based forms for adding, modifying and deleting from the SiteDB table were created. The entry forms are sectioned into the modules that were included in the prototype: Site, Climate, Vegetation, Regionalization, and Soils. Other sections/modules can be added at a later date simply by appending the fields to the end of the SiteDB table and adding a new section to the entry forms. The modular approach breaks the information down into more easily digested bits.

Site Comparisons:
A set of web pages exists for selecting and viewing information from SiteDB. Site information may be browsed in general by site, but parameters may also be selected and retrieved from all or a subset of sites. This ability to compare parameters across sites demonstrates the value of a SiteDB database. The output can be either a listing of one site at a time, or a summary of sites in an html table.

Moderated Database:
In order to insure that updates can be made whenever convenient for the site, an additional design feature is under consideration for the SiteDB prototype. In general it is important to have a database moderated by a single individual or small group. To maintain both the site modification capability and the moderated approach, a temporary table can be created to hold all database updates input by the sites. After reviewing the updates for errors, the moderator can update the main table from the temporary table. Although SiteDB is designed to contain information that rarely changes, sites do need to have the option to modify information pertaining to their individual site.

The Prototype:
SiteDB Main Menu
Site View
Climate View,
Vegetation View,
Regionalization View
Soils/Sediments View
Add Site
Select Site to Modify
Modify Site
Comparing Site Information

Important Goals:
-Network: Single location for information on all sites.
-Views: Organized as basic and topic information -Modular: Flexibility to add new modules.
-URL: Single location for maintaining important site URLS.
-Updateable: Moderated design but easily updateable by individual sites.
-Searchable: Single or multiple sites for all or selected parameters.
-Output: Viewed or exported file.

2.3 Climate Database Project (ClimDB) Update (D. Henshaw (AND))

ClimDB Overview
As part of the LTER Information Managers' Network Information System (NIS) development, the LTER climate database project (ClimDB) has developed a system for harvesting daily climate data from LTER sites (Henshaw et al. 1997). Individual LTER sites still control and maintain all locally collected climate data. Individual sites "filter" data from local information management systems into a standardized exchange format. Data is then harvested in this

standardized exchange format and is stored in a centralized relational database. Applications have been developed to generate two monthly distribution report formats (Bledsoe et al.1996) using this centralized database of daily values. Additionally, a webpage (http://www.limnology.wisc.edu/climdb.html) has been created to provide access to the daily and monthly climate data as well as to permit query by LTER site, weather station, and date, and to provide graphics capability.

Update
A meeting of climatologists, modelers and data managers met at Sevilleta, NM, on October 2-5, 1997 (CLIMSTAN). A meeting combining scientists, data managers and data users proved to be very successful, and should be used as a model for projects involving intersite exchange of data. The meeting report, which includes instructions and requirements for participation in ClimDB, can be found at (http://www.unc.edu/~greenlan/CLIMSTAN.html/Default.htm)

The meeting was successful in accomplishing the following main objectives:

1. Revise the LTER meteorological observation standards
2. Refine and expand the ClimDB exchange format to include more climate variables at more temporal scales, and develop variable naming conventions for this expanded implementation.
3. Establish metadata content requirements for meteorological observations

Robin Stubbs has done all the ClimDB prototype programming including recent revisions to the exchange format and implementation of the new variable naming conventions. Robin is also taking responsibility for transferring the prototype from NTL to the Network Office.

Darrell Blodgett has written a prototype for entering and viewing climate metadata. See (http://www.lter.alaska.edu/msql/w3-msql/climdb/climdb_menu.html)

ClimDB Working Group

Don Henshaw
Robin Stubbs
Karen Baker
Barbara Benson
Darrell Blodgett
John Porter
Emery Boose
James Brunt

ClimDB Action Items

1. Establish ClimDB+/ClimDBplus mailing list (This has been done. Members might be data managers or climate data investigators.
2. Expand ClimDB to include all LTER sites. Only 15 of 21 sites are currently represented.
3. Expand ClimDB data to include all years of available data for participating climate stations. Current implementation had requested only years 1991-1995.
4. Expand variables to include Levels 2 and 3 (See the CLIMSTAN meeting report for details).
5. Have all participating sites populate the climate metadata prototype.
6. Move the ClimDB prototype to Network Office. NTL currently maintains the prototype in Oracle, but the Network implementation will be in MS SQLServer. Robin Stubbs (NTL) is handling this transfer.
7. Other items
    a. Accommodate X-Roots by including soil variables
    b. Include high temporal resolution data
    c. Include secondary meteorological stations

Requirements for Site Participation

Each participating site will:

1) Write a filter to restructure local site climate data into a standardized daily exchange format . This filtering process can occur on a scheduled basis into static files, or can be created dynamically during the harvest process.

2) Provide an Internet address (URL) to identify the location of the exchange format data file. The address will link to a static file or a dynamic script.

3) Provide overall LTER site level metadata, metadata for every weather station, and metadata for every parameter measured at each station.

4) Provide appropriate quality assurance parameters for central database validation checking. Note that these parameters are used to detect errors in transmission or errors in aggregation that may occur at the central harvesting site. Primary quality assurance checks are to be performed at each individual site.

References

Bledsoe, C., J. Hastings, and R. Nottrott. 1996. Xclimate workshop, Davis, California, USA [Online]. Available: http://www.lternet.edu/documents/Reports/Xroots/aclim.htm [1998, November 16].

Henshaw, D. L., M. Stubbs, B. J. Benson, K. Baker, D. Blodgett, and J. H. Porter. 1997. Climate database project: a strategy for improving information access across research sites. In Proceedings of the Data and Information Management in the Ecological Sciences Workshop. Albuquerque, New Mexico, USA [Online]. Available: http://www.lternet.edu/ecoinformatics/guide/frame.htm [1998, November 16]

2.4 All-Site Bibliography Revision (J. Brunt (NET))

A working group has been established to oversee the update and revision of the LTER all-site bibliography (ASBIB). First steps of the group include the updating of the current ASBIB information that is in Harvey Chinn's original WAIS indexed format. This will provide a delimited file of the complete all-site bibliography to work with in developing a new, more functional database. The working group includes: Karen Baker (PAL), Darrell Blodgett (BNZ), James Brunt (NET), Hap Garritt (PIE), Don Henshaw (AND), Peter McCartney (CAP), Eda Melendez (LUQ), and Ken Ramsey (JRN). As a prototype, the group will modify a data model for MS SQLServer that Peter McCartney developed for the Central Arizona LTER. The current bibliography file will be migrated to this database at the network office along with some access and update functions. It is expected that this initial testing will be available for review by early 1999, with a prototype available for presentation at the next annual Data Management Committee meeting.

3.0 Network Office Report (J. Brunt (NET) and R. Waide (NET))

The Network Office has survived the transition from Seattle to Albuquerque (Network News 11:1 Spring 1998) with a few unavoidable side effects resulting from hardware and network changes. Things should be running smoothly now that the office is fully staffed. Colin Johnson, User Support Analyst III and Richard Dahringer, Programmer were hired in May to maintain and improve the LTERnet computer and software environment. They report to James Brunt. With these hires, the helper@lternet.edu mail line is again being monitored on a daily basis.

Improvements are being made to the personnel database to make it easier to maintain and to address more issues that

will help the Network Office support the LTER program. This effort will require help from the Data Management Committee. In addition, the office will be putting forth effort to transition the Network Information System prototypes into running modules at the Network Office. The DTOC module has already been done with the help of John Porter, and arrangements have been made for Robin Stubbs to be able to work on migrating the ClimDB module. It is envisioned that this model of having site information management specialists work on network projects will be continued, as funds become available. In addition, the office will be moving ahead to update the all-site bibliography.

Finally, effort from the committee will be necessary in preparation for the LTER National Advisory Board meeting to be held later this year. This group of prominent scientists will evaluate the LTER program and make recommendations for its future. Information Management will of-course play an important role in the future of LTER as will the Data Management Committee.

4.0 Special Reports

4.1 Internal Organization - Knowledge Management

(C.Wasser (SGS) and K. Baker (PAL))

An integrated, scalable system for storing the collective wisdom of the group would be of benefit to the LTER Information Managers. In developing such a system, it is important to recognize the distinction between 'information' which is data endowed with relevance and 'knowledge' which is a high-value structure of inter-related information. Knowledge management (KM) is an integrated approach to identifying, capturing, sharing and evaluating knowledge while a knowledge management system (KMS) is a searchable, shareable and easily retrievable repository of heterogeneous information types.

A knowledge management system recognizes the importance of all the knowledge stored in ours heads by making this knowledge explicit and available in a centralized or network location. Such a system can be a searchable, shareable, and easily retrievable repository of heterogeneous information types that captures the relationships between pieces of information and people. The benefit of a KMS is the reuse of knowledge and leverage of information into other applications, both current and future.

The LTER data managers have developed a suite of organizing strategies including meetings, projects and surveys. At the onset of a new one of these central units of organization, coordination begins with designation of a specific and easily referenced name for the effort. There are reports and mail associated with each unit effort. In fact, the facilitating and organizing of all forms of group communication is of primary importance contributing to the success or failure of any project.

Meetings may be organized with an online web page. Techniques for optimizing meeting organization have been developing for both the LTER central committee (CC) meetings as well as the data manager (DM) meetings. As data managers, we can learn from observing how the LTER CC meetings are organized with an online meeting coordination page. Pre meeting information may include both an online meeting registration, area logistics and an agenda. The posting of registration information prior to the meeting aids in travel co-ordination among participants as well as improved counts for meals and excursions. Having the area logistics easily available permits an individual to find logistics information when they are ready to consider it rather than having to search among mail messages sent earlier. The page provides a place for posting meeting documents, group reports, and site flashes. Further, the addition

of a 'lessons learned' page by the data manager at the end of a meeting provides future organizers with valuable insight. A central page provides a readily available reference for meeting attendees, but also ensures there is central location for everyone's files. Since group projects exist beyond a single meeting, it is valuable to extend the concept of meeting coordination pages to project coordination pages which could be located at the network office given a network system of organization and privileges such that the LTER site data managers can work centrally. A coordination page also provides a location for identifying and posting action items.

The concept of an online central web page works well for coordination of projects as well. Projects to date include metadata, Network Information System (NIS), Climate database (ClimDB), Data Catalog (DTOC), Site Information (SiteDB), Data Policy/Ethics, MODIS Land Science Team & Long-Term Ecological Research Network Synthesis (Modlers) including Net Primary Production (NPP) and Leaf Area Index (LAI), Northwest Alliance for Computational Science and Engineering (NACSE) collaboration, and Nitrogen Deposition and Dynamics (NDD).

Surveys provide an effective method for gathering information from a group. The subsequent communication of this information electronically to the group through online posting and meeting reports insures that learning proceeds as a group. Surveys to date have included electronic equipment (91-94), soils (93), online data sets (96), bibliographic software (96), software survey (96), data user access (96), collaborations (98), international data management (98) and home page elements (98).

Capturing information for meetings, projects and surveys actually initiates the process of documenting the history of an organization. Just as a librarian is a society's historian so a data manager fills the role of site historian. As such, the establishment of a milestone list is a natural task for the site data manager to oversee. In fact, the home page reflects one aspect of the history of a site so one might consider capturing snap shots of a site's home page by saving versions of the home page html at frequent intervals.

Currently, there are an increasing number of sites with the consequent result of increasingly difficult meetings and projects to coordinate. Having a KMS system in place would capture past knowledge thus eliminating the frustrations of re-inventing past information for each new task. A knowledge management system may be approached on an application-by-application basis, but it is best keep in mind an overall system. Facilitating both access to information and the addition of information is a major determinant of the success of a KM system. Capturing information is often a difficult step so it would be beneficial to identify clearly a process for adding new information as well as insuring relevant links are created. A valuable addition to such a system would be the creation of a searchable index of the information.

Action Items needed from LTER NET for Knowledge Management System (KMS):

1. Consider in conjunction with data managers how to create an IM KMS
2. identifying most important elements of IM KMS to implement:
    a. Develop and post KMS how to's.
    b. Designate disk space for KM document storage (at minimum 50MB).
    c. Designate web space for project level digital meeting rooms and links.
    d. Link project/meeting notes to main LTER web pages.
    e. Provide project leaders effective process for updating and adding to project/meeting pages.
3. Consider how to how to host web page for the next data manager meeting.
    a. agenda
    b. online registration
    c. ongoing registration information summary d. pre-meeting notes
    d. post-meeting notes
    e. final report
    f. lessons learned
4. Address and post specifics of how to get working groups together.

4.2 eXtensible Markup Language and LTER Data Management (P. McCartney (CAP))

What is XML?

Extensible Markup Language (XML) is a language for describing the structure of data. More specifically, it is a *metalanguage* – a language with which one defines a language. One can use XML to create a specialized language (ok…jargon) to identify the various elements that comprise a certain domain of information and define a set of rules as to how these elements relate to each other.

XML is often described as the successor to HyperText Markup Language (HTML), the language of the World Wide Web. While this may indeed one day come to pass, we needn't presume to be fortune tellers in order to recognize some very obvious relations between HTML and XML. First, both languages are derived from a common parental language – Standard Generalized Markup Language (SGML). As such, they both consist primarily of a series of inline tags with optional attributes by which information is appended to sections of text. XML is a subset of SGML, lacking some its complexity and features that are not relevant for Internet. Both languages were designed with the World Wide Web in mind. As with HTML, the primary body responsible for defining and maintaining standards is the international W^3 Consortium. XML is thus a public, non-proprietary language.

What is different about XML from HTML, however, far outdistances what is similar. In HTML, there are only finite number of tags, a list that is independently defined by the W3 consortium. In XML, users of the language define their own tags and attributes. The list of tags can be limitless and custom-tailored to the specific application. In HTML, the information contained in the tags describes only how the text is to be rendered on the screen – it conveys no information about the content, or meaning, of the text. In contrast, XML tags are used exclusively to describe the content, or structure of the data. An XML document cannot be rendered for display by a browser as can HTML – it requires the addition of a style sheet to provide display instructions for each element type in the document. Style sheets may be written as Cascading Style Sheets (CSS) or with eXtensible Style Language (XSL), a complimentary style sheet language developed by the W3.

XML was developed to overcome some specific shortcomings of HTML by provide content description (metadata) for web-accessible documents. Everyone has experienced queries using popular web search engines that yielded long lists of documents, many of which were far afield from what seemed at the time to be a fairly unambiguous search term. XML allows web searches to go beyond simple string matches to more powerful boolean –like queries of the form "WHERE <tag> = search term"

XML also enhances the value of documents by embedding metadata within them. Metadata tags enable automated processing of web documents through the use of a common language that is abstracted form any hardware or software environment. The exchange and migration of Primary intended use is for documents with high, lasting value

Using XML

The first step in using XML is to create an XML document. As with HTML, an XML documents can be created and tagged entirely with an ASCII text editor or through the use of an editor designed specifically to handle XML documents such as Microsoft's XML Notepad. More importantly, XML documents can also be created as output from other programs such as a database application.

To make use of an XML document, it must first be read by a parser, a program that reads the tag structure and then returns the content in some manageable form, typically as a tree hierarchy with the elements and sub elements forming the nodes. Such a data model is then easily navigated to extract the information and perform whatever task is desired. This might be to simply render the data as HTML for viewing, search the data for content to be returned from a query, or insert the data into a waiting database.

It should come as no shock that XML is already in wide use. The most popular application is Microsoft's Channel

Definition Format – yes, that annoying little feature that you so disdainfully removed from your NT desktop. Resource Description Framework (RDF) is another XML based language that has been developed specifically for encoding metadata. Other metadata languages based on SGML include EXXX, a language for archival catalogs, the Dublin Core Metadata standard for libraries and the Federal Geographic Data Committee contend standard for spatial metadata.

Features of XML

The language itself consists of six basic components. These are: Elements, Entity references, Comments, Processing Instructions, Marked Sections, Document type declarations. To quickly grasp the essence of XML, it is only necessary to consider the first and last of these.

Elements are the tag pairs that envelope some data – e.g., <tagname>content</tagname>. The tags may contain attributes in the form <tagname attribute=value>. Elements may include other elements by nesting tags. Some elements may be optional, others required, and still others may repeat. As a simple example, consider how we might mark up content from the LTER personnel directory:

```
<scientist active=true>
<name>
<firstname>Peter</firstname>
<lastname>McCartney</lastname>
</name>
<address>
<department>Anthropology</department>
<institution>ASU</institution>
<state>Arizona</state>
<zip-code>85287</zip-code>
</address>
<corearea>data management</corearea>
<corearea>land use</corearea>
</scientist>
```

Here we have a scientist element comprised of a name, address and core area elements. Name and address are in turn comprised of other elements. An attribute (active) has been added to the scientist tag to indicate that the record refers to an individual currently active in LTER (well, currently on the payroll, at least). Note that the opening and closing tags nest perfectly – this is a requirement of XML. A document with improper nesting is not "well formed" and can not be read by an XML parser.

The rules for governing what elements can be used and how they are nested are specified in the Document Type Declaration (DTD). This component consists of a series of declarations for the various elements, attributes, entities, and notations used in the document. For our sample document above, we might include the following DTD:

```
<!ELEMENT scientist (name, address,corearea+)>
<!ELEMENT name (firstname,initial?,lastname)>
<!ELEMENT firstname (#PCDATA)>
<!ELEMENT corearea (#PCDATA |Populations|Primary Productivity|Disturbance|Data management|Land
Use)>
<!ENTITY lter logo SYSTEM "/images/lterlogo.gif NDATA GIF87a">
<!NOTATION GIF87A SYSTEM "GIF" >
```

Here we explicitly define the *scientist* element as a name, address and one or more *coreareas*. Name in turn consists of a *firstname*, optional *initial* and *lastname*. For *corearea* , we define a domain of possible values it can assume. A DTD is not mandatory but its presence permits validation of the structure. Validation ensures that the content of the document is compatible for whatever processing is to be done with it – for example, it ensures that bibliographic information will be compatible with the database into which it will be inserted.

Application to LTER

We can now look at the role XML might play in the overall data management program of LTER. The most obvious application is as a medium for exchanging data between sites and the NET office. The Network Information System (NIS) project of the LTER data managers has initiated several pilot studies to develop solutions for passing information from site databases to NET. These include bibliographic data, personnel data, site description data and even monitoring data. In many of these cases, solutions based on structured ASCII files and or web based document management tools have been implemented. XML might provide a universal language for creating data exchange files that takes advantage of extensive industry support and development.

Perhaps down the road, but possible as early as the next data manager meeting, plans may be laid for creating an LTER metadata content standard using. The advantages it would offer are standardization across LTER sites with greatly facilitated exchange of metadata and ultimately data. While large data files themselves will likely continue to be exchanged in proprietary formats as with CLIMDB, an XML based metadata standard will permit automated retrieval and processing of those files.

Figure 2 depicts a model of an XML-based data exchange system. In this model, metadata are maintained at LTER sites in a variety of forms such as text, html, or structured databases. Software is used to generate XML documents from these sources which can be transmitted across the web to the NET office, or some other destination. These documents are read by a parser and then further processed as required for one or more applications such as display or data harvesting.

Goals for LTER

In order to realize these future goals, LTER data managers need to begin thinking about three areas of research:

1. Definition of a metadata standard and implementation using XML.
2. Development of tools for creating XML documents from existing sources such as site databases, web pages, etc.
3. Development of necessary parsing and presentation tools for web and data applications

Definition of a metadata standard should include consideration of existing efforts using XML and SGML, in addition to a comparative analysis of LTER site systems. One of these is Resource Definition Framework (RDF), a generic language for developing machine-readable metadata. Another is the Federal Geographic Data Committee (FGDC) metadata standard which has extensive government support. IBM and Microsoft have submitted a draft Vector Markup Language (VPL) which might be used to encode spatial information. Finally, efforts at Oak Ridge National Labs (ORNL) to create a metadata standard based on the FLED paper need to be reviewed.

Mainstream software support for easy creation of XML documents is still months away, but press releases indicate that products like MS Office will include native XML support in one form or another. For generating XML from existing data sources, customized solutions will be needed to accommodate the hardware and software environments at different sites. This will involve some effort from both Site DMs as well as NET personnel in coordinating the necessary software development.

Tools for parsing XML are already readily available with expanded XML support expected in the next crop of Internet browsers and development languages. The point where custom software will be needed for making use of XML documents lies in the interface between the content that is returned from the parser application and the target environment such as an existing database. For example, the pilot XML harvesting application developed for the all sites bibliography consists of some simple Visual basic code to traverse the content map returned from Internet Explorer's XML parser and copy the appropriate fields into a database recordset for insert.

Summary

XML is a language for describing the content of a document. It can be used to define strict rules for the structure of a document permitting automated systems to be better informed about the content of the data they are processing. It provides a non-proprietary way to transfer information across the internet and by virtue of extensive industry support,

can be used to produce standardized output that can be used in quite different applications such as display, data exchange, and search engines.

Sources

- World Wide Web Consortium (W3C) XML Standard http://www.w3.org/XML/
- 20 Questions on XML http://builder.cnet.com/Authoring/Xml20/index.html
- Microsoft XML Web Site http://www.microsoft.com/workshop/c-frame.htm#/xml/default.asp
- Enigma XML site http://www.enigmainc.com/xml/
- Robin Cover's XML page http://www.oasis-open.org/cover/xml.html

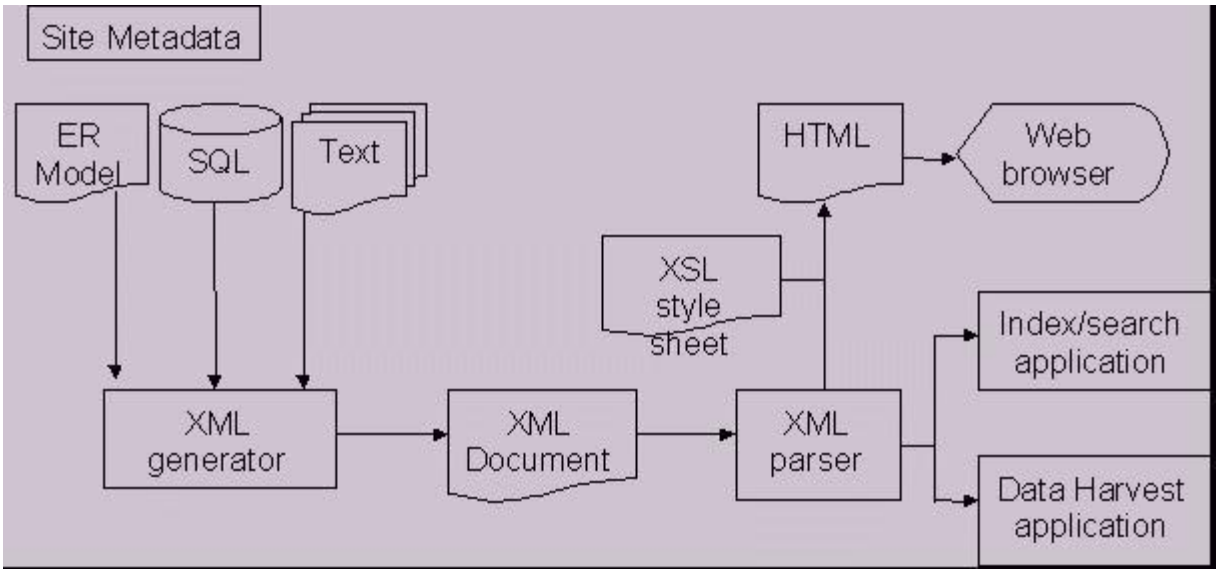

Figure 2. A Model of an LTER metadata system based on XML

4.3 Web Publishing Issues (D. Steigerwald (MCM))

In the process of getting metadata on-line for the McMurdo LTER web site, many of the PIs (principal investigators) were referring Denise Steigerwald (the data manager) to publications that resulted from their research. In order to make it easier for Denise, both at the time, and in the future, computer files were created containing the text from published articles. In doing so, it occurred to her that it may be helpful to provide these files to web users, but that raised some concerns and questions regarding copyrights. Other LTER data managers were contacted for feedback. Responses included the following:

1. Chris Wasser (Shortgrass Steppe LTER): He had decided to just post the abstracts of relevant articles. This way you don't have to deal with the whole article, but it gives more information than just a title. Copyright laws may still apply, but as an educational facility, he was unsure of the ins and outs regarding copyrights.

2. Peter McCartney (Central Arizona / Phoenix Urban LTER): Basic issues are raised about web publishing through this whole process. A related issue is: if we post copies of publications (or even post original material) on our website, what commitment are we making to maintaining the document? Some journals are developing citation styles for internet documents. If people start linking or citing these materials, we will be committing ourselves to maintaining those links.

Assuming we obtain permission to post an article, a very bold disclaimer should be included, saying it is a reproduction for convenience, and any citation should refer to the published, rather than the web document. Otherwise

you may have a headache down the road keeping all those document URLs stable.

Taking into consideration the above issues, Peter had decided to maintain bibliographic references for metadata, and use boiler-plate language from publications to fill in portions of metadata such as research design, sampling etc.

3. Eda Melendez-Colom (Luquillo Experimental Forest LTER): She decided to provide only the abstract on the web. Users can find out what the publication is about this way. If they are interested in learning more, they can request the article.

4. Karen Baker (Palmer Station Antarctica LTER): She pointed out that journals and books all have different requirements, so they must be considered individually. Online articles are useful in many cases. She has made abstracts the expected entry with a bibliographic reference; selected individual articles are provided online with the original publication cited at the top.

Related to the issue of posting articles resulting from long-term ecological research, Denise needed to resolve how to deal with articles ABOUT long-term ecological research. In February, 1998, a New York Times article was published on the LTER research occurring in the McMurdo Dry Valleys. One of the principal investigators asked Denise if she could obtain permission and post this article on the McMurdo LTER web site. Denise therefore contacted the author to see if she could post the article. The author referred her to the editor of the New York Times web site. He granted approval to post the article on the web, but it was no longer available on their site. A scanner could be used to get the article in the McMurdo LTER system, or the text could be typed in by hand. Since the scanner did not provide a clear enough image, the latter option was used.

Shortly after the New York Times article was posted, Denise was informed of a Science article that dealt with McMurdo LTER research. She therefore contacted their editor to seek permission to post the article, and was told that 6 months must pass before they'll consider allowing a link or posting an article. Charges are applied to subscribers, so posting an article on another web site would prevent this from occurring.

Through the process of trying to post the New York Times and Science articles on the McMurdo LTER web site, Denise discovered how difficult it would be to post other full articles containing copyrights. She therefore decided to just provide abstracts of articles that contain a copyright. Complete text articles from journals where copyright is not a problem (e.g., Antarctic Journal of the U.S.) would also be made available.

Further discussion regarding copyrights and posting of articles that result from long-term ecological research followed.

4.4 Publishing Ecological Data (D. Olson (ORNL))

Changes in publication policies of the Ecological Society of America (ESA) were discussed by Dick Olson in terms of the impact on data management. ESA has adopted a policy to create an archive (Ecological Archives) and established a mechanism to publish peer-reviewed data papers equivalent to other ESA open-literature publications (see Peet, R.K. 1998. ESA Journals: Evolution and Revolution, Bulletin of the Ecological Society of America 79(3): 177-181). The publication procedure is being finalized in terms of instructions to authors and the peer-review criteria. Contributions will be reviewed in terms of ecological value and completeness of documentation, and data papers may be rejected if they do not meet the criteria. It is anticipated that data papers will be cited in any and all papers that use the data. Furthermore, such citations will show up in a citation index search and will be counted similar to citations from any other papers published by that scientist.

This venue will both preserve ecological data for future scientific study and provide incentives for individuals to contribute their data to an archive. In additional to publishing data papers, the ESA Ecological Archives will store in digital form most appendices and supplements from ESA journals. The materials may be tabular data, photos, remote

sensing scenes, GIS coverages, video and audio clips, etc. Submissions will be given an accession number that will be included in the citation format. Currently the archive is hosted by the San Diego Supercomputer Center; however the URL will be published in each issue of the various ESA journals to provide an accurate pointer to the data archive.

4.5 DAAC Web-based Information System (D. Olson (ORNL))

Implementing a distributed, Web-based data system for the LTER Network was discussed during the workshop in terms of harvesting metadata from individual sites and providing a central search and order capability with keyword and free text searching. Dick Olson described the Mercury project at the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC http://www-eosdis.ornl.gov/). Mercury is a Web-based, distributed system designed to allow the searching of metadata to identify datasets of interest and to deliver those datasets to the user. The goals of the system are to facilitate early exchange of data between PIs, provide PIs complete control of data visibility in the system, achieve rapid and economical deployment, and incorporate automation and scalability to reduce maintenance costs.

The system software is centralized using specially modified commercial-off-the-shelf software components. The metadata files and data files reside on data-provider machines distributed anywhere and connected by the Internet. Data providers make their metadata and data available to the system by simply placing them in a 'visible' area on their machines, and periodically the Mercury crawler harvests the metadata and automatically builds both an index and a relational database with the information, which resides at the central facility. Initial results have been encouraging to provide metadata harvesting from distributed sites, search functionality, and capabilities to convert user-defined metadata into standards such as FGDC and Z39.50.

4.6 Office/PI Status Survey (B. Benson (NTL))

Barbara Benson reported briefly on the results of an informal email survey she had conducted in April 1998. Data managers were asked whether they had PI status and whether they had a private office. Twenty sites responded to the survey. At sixteen sites the data manager had a private office. At 7 sites the data manager had PI status.

5.0 Working Group Reports

5.1 NIS working group report (P. McCartney (CAP))

The NIS working group opened with a general discussion of what everyone understood NIS to be about and what issues concerning its future directions and long-term maintenance have arisen from the pilot projects implemented thus far. Discussion items raised included:

- **Scope of NIS**. The current components of NIS were identified as the All-sites Bibliography, SiteDB, the Data

Table of Contents (DTOC), the all-sites personnel directory (PersDB), and the climate database (ClimDB). With the exception of ClimDB, most share a common theme of providing metadata-level information about sites and their research. Questions were raised about how should these databases be related – should they remain discrete systems, or should they be integrated with linkages or even jointly managed in a common database?

- **The role of NET in the development and maintenance of NIS**. Should NET provide central storage and access for NIS databases? Should NET personnel have more active role in development or should their role be to implement and maintain database solutions developed by the pilot teams? How do these questions affect distribution of resources? Are there significant technology transfer issues associated with migrating solutions developed as pilots by one site to NET?
- **Redundancy**. Do the data services provided through NIS represent a unique product that adds value to the databases maintained by the sites or does it simply represent duplication of information? In other words, will continued development of some NIS components (such as the all sites bibliography) eventually eliminate the need for individual sites to maintain their own site-specific databases? It was recognized that the most costly solutions involve maintenance of both local databases and a central integrated one. Less costly solutions include either abandoning the local databases in favor of a single centralized database or developing centralized search tools that can query the local databases directly without requiring maintenance of a central copy.
- **Communication**. Some individuals expressed the desire to be better informed about NIS and the progress of individual components. Suggestions included a dedicated web site at NET where documents on NIS could be centrally posted and maintained by participants rather than the current situation of NIS-related information scattered among the webs of those sites conducting the individual pilot projects.
- **Technology**. It was recognized that the current list of pilot projects represents a diverse array of technologies and overall designs. Distinctions were made between "deep" databases that contain many records with tightly controlled fields (e.g. ClimDB) and "wide" databases with fewer cases but greater diversity in their structure. Pilot projects seeking to build central compilations of data from site databases (ClimDB, All-sites Bib, PersDB) were contrasted with those seeking to merely build central indexes of site databases (DTOC). Finally, the need to separate the design of these project from their underlying technology was discussed to facilitate transfer of these components (to NET, for example) as well as insulate them from the inevitable march of technological advance and obsolescence.

Based on the above discussions, two goals were identified for the NIS group to be accomplished during the following year:

1. Review existing documents on the goals and design of NIS and its components. Specifically, we need to reaffirm the need for that component, the target audience, and consider the issue of redundancy with site products. It was proposed that the NIS team would seek to hold a meeting to consider the results of this review and to draft a new long-term plan for NIS with a scope of at least 5 years, possibly longer.
2. A comparative review of the technologies that have been applied in the various pilot studies that have been carried over the last few years for NIS components to identify elements of technology that might be shared between these efforts or applied toward new initiatives within the NIS program. This review should focus on breaking down the various solutions into discrete procedural and technical components that would be transferable and reusable in future projects. While the venue and deliverable product for this action was not identified in the meeting, it would be reasonable to expect that it could be accomplished through an email survey and a web-based summary of results would be produced by the working group leader (McCartney).

Current NIS Working Group

Karen Baker
Barbara Benson
Darrell Blodgett
James Brunt
Ned Gardiner
Don Henshaw
Peter McCartney
Eda Melendez-Colom

John Porter
Ken Ramsey
Robin Stubbs

5.2 Data Ethics (J. Briggs (KNZ) and E. Boose (HFR))

At the 1990 IM meeting, the Data Managers formulated a general data access policy that could be modified and adapted for use by individual LTER sites. At the 1997 IM meeting, with encouragement from NSF, the Data Managers simplified the earlier policy (reducing the number of data types from four to two) and recommended that the new version be adopted as the official LTER Network policy. In October 1997 the updated policy was presented to the LTER CC meeting and accepted by vote. In February 1998 it was presented to NSF at the LTER Exec meeting. NSF endorsed the new policy, indicating that such a policy is essential to the leadership role played by the LTER Network in the ecological community, especially in the area of data management.

At the October 1997 CC meeting, concerns were expressed by the PI's about the possible unethical use of online data. The Data Managers were asked to follow up on one of the recommendations from the 1997 IM meeting: contacting professional societies such as the ESA to help them develop a code of ethics for the use of electronic data.

At present the ESA does not have an explicit statement on the ethical use of online data. However the ESA's Code of Ethics does contain a general statement (no. 2) on the use of data for publications: "Will not submit for publication any manuscript containing data they are not authorized to use. Where not otherwise specified by contract or explicit agreement, the principal investigator(s) of a research project should be assumed to retain the right to control authorization or use of the data" (Bulletin 1996, 77(1):22).

At the 1998 IM meeting, the question of online data use was addressed by a working group as well as by the entire committee. It was recommended that J. Gosz draft a letter to the ESA and other professional societies informing them of the LTER Network data access policy and encouraging them to formulate an explicit statement on the ethical use of online data. In particular, the letter should make the following points:

(1) In the future more and more scientific data will be made freely available to the scientific community through the use of online data archives. The LTER Network is a good example of this trend.

(2) Such data may be used for publication only with the permission of the investigator and with proper citation.

(3) In particular, users of online data must follow the specific data use policies (if any) that are presented as part of the online data archive. Such policies might appear, for example, in a "Data License" on the site's web page.

(4) Professional societies should develop specific guidelines for proper citation of online data and metadata.

5.3 LTER 2000, Outreach, MSI Revisited, Y2K Working Group (S. Stafford (AND))

Group members: Garrett Ponciroli (KBS), Gregg MacKeigan (SEV), Chris Wasser (SGS), Bob Waide (NET), Miguel Acevedo (ILTER), Mike Hartman (NWT), John Helfrich (ARC), and Susan Stafford (AND).

MSI Revisited: The timelines of the MSI (Minimum Standard Installation) has come and long gone! Gregg

MacKeigan will draft a revised MSI that addresses the issue of e-mail attachments.

Outreach: Outreach from the LTER network can be viewed in the terms of both data and expertise.

LTER 2000: There is an opportunity to showcase the contributions IM has made to long-term ecological research at the Aug. 2-5, 2000 LTER 2000 meeting. These will include: (1) The preservation of data, (2) promotion of metadata, (3) facilitation of synthesis research across sites, (4) promoting the issues of data access, (5) conversion of data into information into knowledge, and (6) the overall mission and vision and new role of Information Management and the synergy between IM and research/scientific inquiry.

We want to demonstrate the user of LTER data by others. For example, using LTER data to calibrate global models. We could go back to the people who have used the LTER data and create the "poster children of LTER". We would like to showcase the working model of NIS for example "Nitrogen" across the NET.

Y2K Compliance:

Web pages on the Year 2000 compliance problem have been established at (http://www.nsf.gov/oirm/y2k/start.htm) and (http://jornada.nmsu.edu/site/year2000/year2000.htm).

6.0 NACSE Collaboration and Pilot Project Proposals (D. Henshaw (AND))

The Northwest Alliance for Computational Science and Engineering (NACSE) promotes the development of highly interactive Web-based resources (http://www.nacse.org). The LTER community provides an organized testbed for early deployment of NACSE projects associated with Web access to large-scale scientific databases. Ongoing collaborations between NACSE and LTER sites were initiated in February 1998 at a DataTask Meeting in Corvallis, Oregon. Web pages describing this collaboration (http://www.nacse.org/lter.html) and meeting notes and discussion have been established (http://www.crseo.ucsb.edu:80/lter/dm/projects/nacse/).

NACSE has strong computer science expertise and wishes to partner with research groups like the LTER. NACSE is looking for proposals on which more formal partnerships can be established. Generally, they are looking for proposals that are interdisciplinary and cross-site oriented. They can be site specific as long as they are scalable. The following criteria has been requested for inclusion in these proposals:

Criteria for NACSE:

1. Queries will need to cross multiple databases (multidisciplinary).
2. Why will this enable new science?
3. What is the status of any database(s) being considered?
4. What is the database software being used?
5. What are the approximate number of records?
6. How populated is the database? [How representative is the data? - some of "all" types is BETTER than ALL of one type and none of rest.]
7. Are there novel kinds of access, e.g., visualization?

Subsequent to the 1998 IM Meeting, the following four proposals were developed and submitted to NACSE from LTER Information Managers: Chris Wasser (SGS), Ned Gardiner (CWT), Robin Stubbs (NTL), and Peter McCartney (CAP).

6.1 Development of a Tool for Graphical Display of SQL Data via the Web (C. Wasser (SGS))

Proposal to the Northwest Alliance for Computational Science and Engineering

August 28th, 1998

Introduction

The last few years have seen incredible improvements in the ability to serve database information to Internet users via web browsers. This development has already had significant impacts on the availability and use of ecological datasets. Sharing ecological data between researchers and projects has lead to several new and important avenues of collaborative research that would have been much more difficult only a few years ago. We feel the next logical step in this progression would be the development of web-based tools for on-line interpretation of these data. Specifically, graphical presentation of data would allow users to perform preliminary analyses through their browser.

Typically a researcher will download data and use any of several commercial packages at his/her desk to analyze these data. This analysis can range from cursory overviews, "data fishing," trend analysis, or sophisticated statistical analyses. We contend that the ability to conduct these first three levels of analyses on-line will produce more efficient and effective use of datasets available over the Internet. By conducting these analyses on-line, a researcher can determine at a glance whether or not the dataset is relevant to their research questions and if it makes sense to download the file(s).

In addition, we feel that a tool like this would also enable easier access to graphing tools than would otherwise be possible. A non-commercial alternative to expensive web-enabled graphing packages would allow many education and non-profit researchers to take advantage of the power of graphical presentation in the familiar, easy-to-use setting of a web browser.

Proposal

We propose that NACSE develop a tool for creating graphical representations of data from SQL compliant databases. Such a tool would run through a user's browser and would allow for user input of variables. We envision that such a tool would include, but not be limited to the following functions:

- Simple line graphs: these would probably receive the most use for data exploration and cursory explorations
- Simple column graphs
- XY Scatter plots
- Stacked column graphs
- Multivariable line or column graphs

Such a tool would give a researcher the ability to find "holes" in the data and determine if it is useful for their purposes. A graphical presentation of the data would make these deficiencies in the data much more apparent and save valuable time. Such a tool would also allow researchers to conduct an initial level of pattern or trend analysis to get a rudimentary feel for the dataset. Again, a graphical presentation would highlight these trends much more clearly than examining the numbers. Lastly, this tool would give users an easy way to spot outliers or other data anomalies quickly and easily.

Test Databases

Eventually such a tool would be useful to researchers beyond the realm of LTER or ecology, but we would like to test the tools on several LTER site databases. Initially, the Shortgrass Steppe and North Temperate Lakes LTER databases could be used as test systems. These site both store their data in SQL compliant databases (Oracle) and provide access via the Internet. Both of these databases contain many years of data and tens of thousands of records. The use of these two systems would allow testing of these tools on a wide variety of topics ranging from climate data, to net primary productivity, to limnological data. This broad range of topics will be valuable to ensure that the proposed graphing tools are generic enough to meet the needs of many different disciplines.

6.2 Integrated Data Visualization for Long-term Regional Ecological Studies (N. Gardiner (CWT))

Proposal to the Northwest Alliance for Computational Science and Engineering

Background

A recent initiative by the National Science Foundation (NSF) broadened the geographical scope of Long Term Ecological Research (LTER) research at 14 of 21 sites. These sites are known as "regionalization" sites within LTER. At Coweeta, this broadening of scope has challenged researchers to coordinate across disciplinary lines. Before regional studies began, Coweeta researchers most often worked in areas of known history and land use. Moving outside of the Coweeta Hydrologic Laboratory boundaries required us to integrate socioeconomics, anthropology, forestry, hydrology, biology, and ecology. Since researchers do not control management decisions on private lands, they must understand the many social and physical constraints on decision making and ecosystem function on private lands where research is now conducted. Geographic Information Systems (GIS) play a pivotal role in this integration. At the outset of a project, GIS provide geospatial databases for each individual researcher. Once a project is underway, spatial data underlie synthesis efforts.

Problem Statement

Like other researchers, I have had to redefine the functions and services provided by my office. Before regionalization, this office provided maps and developed data layers for the 1625 ha Coweeta basin. In those days, it was sufficient to work with one or two software vendors to provide these services. Now, we serve data from the entire southern Appalachian region, an area three orders of magnitude larger. Not only has the geographic area increased; the demand to integrate data from multiple disciplines has grown and is of primary importance. These issues demand a return to fundamentals: GIS is a system (software, hardware, and personnel) for storing, analyzing, and retrieving spatial data. Despite public perception, GIS is not merely software. Now we must integrate a full suite of spatial analysis and visualization capabilities: graphical, database, statistical, and geospatial.

I am requesting assistance from the Northwest Alliance for Computational Science and Engineering (NACSE) to develop object-oriented tools for data visualization and analysis. The tools must be as simple as possible for the end-user, who will use a worldwide web (WWW) browser to interact with the programs and who may not be familiar with the software applications- commercial and non-proprietary- which support his analyses. This proposal details a single project, but modular, object-oriented programming will ensure portability, enhancement capability, and technology transfer to other LTER sites.

Proposal

Fish and insects are routinely separated into guilds based on morphological or behavioral traits. We are analyzing land use effects on functional guilds of fish and insects. The following problems are slowing our analytical methods:

1. DBMS. Separating species into guilds is a tedious process without the benefit of database management systems

(DBMS).
2. Database Integration. Once species have been assembled into groups, one must query spatial databases for appropriate land use variables associated with the data points where these species were collected. We must integrate spatial and non-spatial databases.
3. Exploratory Data Analysis. Ecology is the study of complexity. Rarely does one find a univariate factor that has a linear effect on a species or group of species. Multivariate methods are still more problematic. We need tools that perform exploratory statistics (e.g. ensuring normality of dependent variables, correlation analyses, etc.) to guide our empirical and mechanistic model building.
4. Hierarchy and Scaling. Often, factors which are important in explaining behavior or other ecological attributes at one scale are irrelevant or secondary at other scales (larger or smaller grain and/or extent). We need tools that facilitate multi-scale analyses, quickly updating results as larger and smaller areas are considered in association with biodiversity data.
5. Interdisciplinarity. We (Coweeta researchers) have developed a set of explanatory hypotheses important in our biodiversity studies:

    - Historical land use has a prolonged impact reflected in current biodiversity patterns.
    - Riparian geometry is correlated with biodiversity
    - Nutrient enrichment, temperature modification, and organic inputs alter high gradient stream function and structure.

    Other research requires integration with our analyses of stream function and structure.

    - We have developed empirical, spatially-explicit models of land use/land cover transitions in the southern Appalachians based on socioeconomic factors.
    - We are developing a model of carbon allocation to roots and respiration for forest cover.
    - Census surveys reveal long term (pre-settlement, 1800's, early 1900's, and recent) changes in settlement patterns.
    - Sociological research relates peoples' attitudes to their land use decisions and migratory patterns.

    These themes have been investigated individually to this point. We must begin to integrate results and develop hypotheses linking the disciplines represented.

Justification

These obstacles are not going to be solved quickly. In each case described, we have developed our models in isolation of one another. This proposal would serve two purposes. First, it would present results and data to the WWW community. Second, it would facilitate interdisciplinary explanations of stream biodiversity patterns. Publishing data and results on the web with a user interface for analysis will allow researchers to interact with one another's data to create more holistic explanations of ecological structure and function.

Current Status

1. Stream researchers from Coweeta have collected fish and invertebrates from 36 sites throughout the Little Tennessee and French Broad River Basins. We have supplemented our collections with biodiversity data from approximately 50 sites in the Little Tennessee River Basin (courtesy of the Tennessee Valley Authority).
2. Land use data have been processed for 1950, 1970, and the early 1990's for the focal watersheds.
3. Stream classification algorithms are in the final stages of development. Once complete, it will be simple to query stream land use data hierarchically by stream order. This project integrates information about source areas and riparian geometry adjacent to sampling sites.
4. Socioeconomic models of land use transitions have yielded valuable clues about historical land use and areas which are particularly vulnerable to future clearing, exploitation, and disturbance.
5. Census and land holding data from multiple dates are being digitized.
6. Sociological research may yield mappable patterns.

Requested Deliverables

The following is an outline of the desirable functions of the proposed WWW system. Some vendors provide sophisticated integration features (e.g. ESRI's Spatial Database Engine) which perform some of the operations specified. From the end-user's perspective, these solutions are neither economical nor optimized to any particular application. Collaborating with NACSE, I hope to develop a system optimized for ecological research within the LTER regionalization community.

1. A basic template for querying multiple data types
    - Simple user interface
    - Reusable code
    - Template for objects so that vendor-specific methods may be written as need arises
2. Software integration
    - Proprietary vendors
        - ESRI- Arc/Info and ArcView
        - Splus
        - Pathfinder Office (Global Positioning System)
    - Non-proprietary software
        - mSQL
        - gnuplot, octave, et al. plotting and analytical tools
        - SWARM and/or other modeling packages
3. Functions
    - Data visualization
        - Spatial
        - Graphical
    - Statistical Analysis
        - Exploratory
        - Descriptive
    - Modeling
        - Agent-based
        - Compartment-based

6.3 Database / Generic Tool Wish List (M. Stubbs (NTL))

Proposal to the Northwest Alliance for Computational Science and Engineering

1) Batch input to interactive web forms utility

It has become extremely common to encounter a system where web forms have been set up for interactive input into a database. But what if you need to bulk load numerous records from a file into such a system? I haven't seen a good solution to this problem anywhere. I have seen a solution using the perl libwww and I must say it looked pretty awful. There needs to be a higher level solution. This would require serious knowledge of the http protocols. I think someone could achieve global notoriety for a good solution to this problem. It would of course be essential to be able to recover from and log errors from the web form submission so that the operator could tell what records were entered and which failed. This undoubtedly would not be trivial.

2) GIS graphics

We have GIS data, much of which is proprietary. Proprietary interests sometimes prevent letting the public have access

to the actual data files, but allow graphical displays of the data. We would like to have a web interface to the data so that users could pick out some information and have a map image created based on their solutions and displayed to them via the web. (NOT xwindows or some other rare display system.) We are aware of some very pricey commercial solutions to this problem that we cannot afford, e.g., ESRI offers such a product. These could be done from either Solaris or NT server. We would not be putting such large data into a database.

3) Generic web query system

Some LTER sites could use a generic query system to provide intelligent access to their data. The actual mechanism of talking to a particular database would have to be solved on a per case basis and hopefully the database-specific logic could be contained in a database specific driver layer. ODBC might be a way to go with this since numerous databases provide ODBC drivers. JDBC also is a possibility. The essential thing of the generic query system would be the least imaginable customization per table. I am envisioning a table driven system where the table-specific information would be in the selfsame database where the data resides. The key thing would be NOT altering and recompiling the data query program to add a table to it but instead just adding a record to a table.

I would imagine a query control table to have items like the following:

table_name, schema, column list, column types, sort order

table_name: would be the name of the table in the database.

> schema: would be whatever is tacked on before table_name to find the table in that particular database (if anything)
>
> column list: would be a comma-delimited list of the columns that are allowed to appear in a query and in the order in which they should appear
>
> type_list: would be a list of data types for the columns listed in column list, and in the exact same order (e.g., number(3,2), date, char(30))

We would want the query program to:

1. present a list of table_names from the query_table to the web user (the web user picks a table)
2. present information about data policy and possibly have the user fill out a form for access, which is logged
3. assuming 2 is successful, have the user pick fields from a list of the column_names (from column_order)
4. program accesses database to get statistics about the chosen fields and then presents type-appropriate (from column types field) range selections to the user for all the columns (range selection optional default to anything)
    1. program asks user to select sort order (from column list). Do not limit to columns picked by user for output.
5. program asks the user for the type of output wanted: comma-delimited, tab-delimited, html table
6. program asks the user for a limit on the number of data items retrieved
7. program does it.

We currently have data in Oracle. Also, we have NT SQL Server that can be used as a test case.

6.4 LTER NIS: Development of Tools and Procedures for Replicating Content from LTER Site Databases (P. McCartney (CAP))

Proposal to the Northwest Alliance for Computational Science and Engineering

Overview of NIS project

The Network Information System (NIS) is a collaborative project among the 21 LTER sites to provide central access through the LTER network office to information derived from individual site databases. Because of the variability in platforms, structure, connectivity and security of site databases, the more obvious commercial solutions such as replication are not practical at this time. Therefore, a series of pilot studies have been undertaken to explore possible solutions for discreet components of this proposed, shared information system. The purpose of this proposal is to outline a series of technical challenges facing the future expansion of NIS that may warrant collaboration between LTER and NACSE.

NIS Components

Bibliography

The All-Site Bibliography contains bibliographic entries for all literature produced by LTER sites. It compliments the site bibliographies, which tend to provide more extensive coverage for their specific region but vary in their interface and online accessibility. Although an exchange format was defined and a single master bibliography compiled several years ago, there has been no subsequent work done on advancing the system. Current plans involve developing a bibliographic database in MS SQL Server based on the original exchange format. This will be managed at NET with a modified version of the exchange format, perhaps in XML, will be used to manage updates to the system.

Personnel (PersDB)

The personnel database is perhaps the component exhibiting the most redundancy with the individual site databases. While there are clear advantages to site data managers to have personnel directories stored locally, there is a need to replicate these data to a central database so that cross-site collaboration is facilitated. Presently, most personnel entries are duplicated by hand through a web-based entry application, although some sites have executed remote SQL queries to insert batches of new records.

Site descriptions (SiteDB)

A Site Description Database was developed to quickly find general descriptive information about each of the sites. Web based forms interfaced to a relational database were used to enable local site data managers to add, modify and delete both basic and theme site information about their sites in a centrally managed SiteDB database table. Important URL's were also added to the database to provide a single location for recording and maintaining important links to individual site information, i.e., URL for ClimDB harvestable files, other network level projects, etc.

Climate database (ClimDB)

The climate database is the first pilot study to synthesize actual research data. A standardized ASCII exchange format was defined for loading the required fields into an Oracle database. Each site translates their climate data into the standard format and posts in on an accessible internet site for harvest by the project team. Data are maintained by dumping previous contributions and re-harvesting current files from the sites. Newly harvested daily files are summarized monthly and briefly quality assured to prevent transmission or summary errors. Daily and monthly Oracle databases are queriable through a web interface into a variety of report formats and graphs.

Data Table of Contents (DTOC)

The Data Table of Contents was created to enable search access to the dataset catalogs and associated metadata from a central application. A web indexing tool (WebGlimpse) provides the core technology. To overcome the fact that metadata are stored at some sites in databases and are served as dynamic html, an ASCII exchange format providing the title and metadata URL string for each dataset must be generated and posted for harvest by WebGlimpse. The software then retrieves each URL in the catalog and indexes the returned page on the central system. Datasets are located by searching this index and the user is directed to the appropriate web site for the full descriptive information. Updates are achieved by rebuilding the ASCII table of contents file, which is re-indexed regularly.

Strategies

The strategies that have been employed in these pilot projects vary widely. At one end of the spectrum, defined as varying from most to least centralized, a solution which involves the creation, and maintenance of a single, central database is represented by the SiteDB project. In this case, the database is maintained and queried through a central, web-based application rather than being populated from distributed donor databases. Other projects such as BiblioDB, ClimDB and PersonnelDB involve a central database that partially or wholly replicates information from individual site databases. In addition to designing the desired search and reporting functionality, these projects have addressed connectivity with remote databases; selection, exchange, and translation of required data; and issues concerning on-going maintenance of the central system without diverting resources from the original systems. Finally, the least centralized strategy is that taken by the Data Table of Contents, which uses webcrawler technology to harvest only the minimal information to construct an index of the source databases. Basic storage and retrieval functions are not duplicated centrally but rather are performed remotely by the source databases, which are typically collections of html pages.

The best strategy for each component of the NIS will ultimately depend on the nature of the source databases and desired functionality of the centralized product. However, future expansive activities will benefit from comparative examination of these approaches and the development of generic tools and protocols to be used in their implementation.

Identification of Needs

The NIS working group has identified two primary tasks for the 1998/1999 year. One is to revisit the overall goals for the NIS, to better define the purposes and desired functionality for each of its components, and how those relate to those of the source databases maintained by the individual sites. The second goal is to perform a comparative examination of the technologies that have been employed in the development of each component. The hope is that a set of guidelines, protocols, and software tools can be developed that will make it easier to expand and improve the system in the future. Within this second goal, it is possible to identify several problem areas that each pilot project has been forced to address.

Development of standardized exchange formats

Each project has developed its own proprietary exchange format, usually some sort of ASCII file that relies on punctuation or tags to render the data "readable" by the central system. Use of more standardized ways for defining exchange formats would facilitate development and portability of tools for parsing and exporting data exchange files.

Connectivity

Source databases exist on a variety of platforms with varying protocols for interfacing. A suite of customized client tools is needed to connect either natively or via established connectivity protocols such as ODBC/JDBC to the source databases and perform necessary translations, recoding, and reformatting of data to meet the requirements of the central database or search engine.

Synchronization

For systems based on maintaining a central copy of data, or merely a centralized index, there is a need to develop reliable, automated solutions for maintaining that centralized resource through synchronization with the source databases.

Proposed activities

As an interface between the NIS project and NACSE, we propose the several areas as possible foci for specialized information technology projects. While these specific elements may not necessarily work together in a single solution, each represents a significant technological development that may offer great potential toward some the issues raised

above. However, each will require the input of some relatively sophisticated approaches to their application if this potential is to be realized.

Use of eXtensible Markup Language as an exchange specification

XML offers the ability to create data exchange formats that can be read, if not produced, by commercial shrink-wrapped software such as web browsers. The proprietary data exchange formats developed by the different NIS pilot projects could easily be implemented in XML. While XML is not complicated, it is more rigid than HTML and not easily created with simple editing tools. Most sites would require custom tools for translating their database records into XML – a generic translation tool that could be configured by data managers for their databases would minimize the custom programming that would need to be done. Likewise, software to import XML documents into a centralized database would need to be developed.

Develop connectivity solutions using JDBC/ODBC for the various source databases

Another problem requiring technical sophistication is establishing reliable connectivity within a distributed and heterogeneous network. Java Database Connectivity represents a protocol similar to ODBC but capable of working across remote networks and diverse platforms. Development of JDBC drivers and associated software for the various database platforms represented by the LTER would facilitate the development of translation tools.

Develop applications of generalized search engines such as Z39.50 to support searches of multiple databases

Due to the variability in structure of the source databases contributing to NIS, there is a need for software capable of searching and querying diverse database platforms and structures with minimal need for customized client software. Technology such as Z39.50 may be valuable either as an outright front-end for search applications to site databases, or as a translation tool for connecting and harvesting data for the purpose of updating a central SQL database.

7.0 International LTER (ILTER)

7.1 Information Management (IM) in the Latin American LTER

(Miguel F. Acevedo, Member of the IM Committee for the Latin America Network)

Miguel Acevedo

University of North Texas, Denton, Texas 76201
and Universidad de Los Andes, Merida, Venezuela

This presentation was made to familiarize the US LTER Data Managers with the status of the IM activities in the Latin America network and encourage further collaboration in this area.

Seven countries are already part of the network: Argentina, Brasil, Colombia, Costa Rica, Mexico, Venezuela, Uruguay. Some other countries have expressed interest in participating and an increased number is expected (Waide et al. 1998).

Three levels of organization can be recognized: site, national, and regional. The sites associate at the national level, then national offices are associated at the regional level (i.e. Latin America). The increased size and complexity of this network represents a challenge for coordination.

The origins of the Latin America LTER network date to three workshops conducted during 1996-1997 in Puerto Rico, Panama and Brasil (Waide et al. 1998). During December 1997, the US LTER organized a workshop, held in Sevilleta, on Data Management Training for the Latin America network. This meeting was designed to expose the Latin American participants to the Information Management activities in the US LTER. It was a very successful meeting and created the basis for a Latin America working group. A web report was generated during the meeting and posted right at the end of meeting. Issues of metadata, surveys and information manager's profile were discussed. This meeting also served as preparatory for the fourth Latin American Workshop that was planned for 1998.

The 4th Latin American LTER Workshop, convened in Ciudad Guayana, Venezuela, May 30 - June 6, 1998. This meeting was divided in two groups following the US LTER model: Scientific Coordinators and Information Managers.

An effort was made to develop the report in html format during the meeting and post it on the web right at the end of the meeting. Please visit the following URL's:

Taller Puerto Ordaz -Venezuela
http://www.oikos.unam.mx./lalter
http://www.ots.ac.cr/lalter

The following were the three major IM themes discussed in the 4th workshop: Standards, Metadata, Surveys.

Metadata was seen as a first step for exchange. A working group discussed and analyzed several standards (e.g. the one developed in Costa Rica) and recommended the following minimum requirements:

- id code, title, keywords, authors
- context, funding, summary
- sampling area description
- spatial, temporal and taxonomical coverage
- methods, equipment, variables
- language, format
- relationships to other data
- availability, contact, log.

A survey was designed in Sevilleta for assessment of capabilities, status, infrastructure needed and to create a directory. The survey was posted online prior to the 4th workshop, and during the workshop an effort was made to analyze and interpret the results, in order to disseminate these and provide feedback to the field stations.

By the time the 4th workshop convened, survey response was available from the following number of sites by country: Mexico (1), Costa Rica (2) and Venezuela (9). A slightly different survey was conducted in Venezuela and acquired by hard copy, fax, phone. Direct contact with the field stations helped to enhance the response.

The survey provided an assessment of computers, connectivity, data management, databases, data policies, and research projects. These results indicate that the sites have mostly PC's, few dedicated lines and metadata, and some have E-mail capabilities.

How to develop capabilities in IM? in both human resources, computational resources. We need to have plans to improve, connectivity and computers. The importance of human resources was well recognized: profile, relationship to scientists, training.

Other results from the 4th Workshop were: establishment of an IM coordinating group, E-mail group and directory, and decision to organize the next meeting (1999) in Colombia.

Challenges: funding, connectivity, network large size and different languages, striking a balance between national and regional interests, coordination, legacy data. There is a need to sharpen: goals, mission, vision, strategic plans at the level of sites and national networks.

Opportunities: latitudinal transect, scientific cooperation, comparisons, neo-tropical ecosystems, megadiversity areas.

The following URL is temporarily serving as the web site for the Latin American LTER: http://ches.ing.ula.ve/LA-ECO-RED (Taller Puerto Ordaz -Venezuela). From this page you can subscribe to e-mail lists, check for announcements, and check the ongoing pilot project on climate variability.

References:

Waide R., C. French, P. Strott and L. Williams (compilers). 1998. The International Long Term Ecological Research Network. US LTER. 109 pp

7.2 Latin American-ILTER Workshop Notes (Eda C. Melendez (LUQ))

This workshop was held in June 1998 in Ciudad Guayana, Venezuela. The planning for this workshop was realized at Sevilleta in November 1997 with the participation of five Latin American countries (Brazil, Costa Rica, Mexico, Venezuela, and Uruguay). A statement of the mission and the goals for this preliminary workshop are presented at http://www.lternet.edu/~jbrunt/DEC-LA-ABQ.html.

The participants at this workshop produced a draft for the agenda and some preliminary tools such as a survey to appraise the research stations' computer and telecommunication infrastructure, and a "Next Step" outline as a guide to further prepare for the meeting at Venezuela. These are posted at http://lternet.unm.edu/ilter/latin_america/im/.

The spirit that prevailed in the Information Management session of the Venezuelan meeting was one of hard work and clear purpose to assist the scientific community of the Latin American network. Delegates from several stations presented general information about their facilities and research activities, and working group sessions followed. The following is a list of the common needs and issues presented:

1. Standardize methods of presenting data.
2. Have access to a central, coordinating institution that will provide a leadership role in the activities held at the stations, and that will provide them with funds to accomplish them.
3. Develop a criteria to accept research stations as part of the local (country) network which should include the existence of:
    a. Some basic long-term data, such as meteorological data.
    b. A basic computer infrastructure, minimally at the sub-regional level, that will offer computer resources and telecommunication facilities to their research stations.
    c. Human infrastructure (scientists) interested in sharing their data (and information) with other scientists.
4. Establish connectivity at the stations.

These common issues were addressed in the working groups. A document of minimum metadata standards, minimum connectivity standards, and the results of the survey are included in the meeting's Web page (http://ate.oikos.unam.mx/lalter/informacion.htm).

The meeting concluded with general enthusiasm for doing projects as a group. E-mail groups have already been established by the Venezuelan delegate. In general, the working environment and issues presented are similar to issues that USA-LTER Information Management is dealing with. A distinctive characteristic was the press coverage coming at the request of CONICIT, the main sponsor for this activity.

The next meeting is projected to be held at Columbia in the last quarter of 1999.

8.0 Information Management (IM) Committee Organization and Policy (D. Henshaw (AND) and J. Porter (VCR))

8.1 IM Organizational Structure Changes

The IM Committee unanimously adopted the creation of a chairperson position to bring the LTER IM Committee structure in line with the structure of other LTER committees. The chairperson will act as a representative of the entire committee, and will not represent any specific site. The chairperson will either attend the biannual Coordinating Committee (CC) Meetings or assure that the IM Committee is represented. The term will be three years and will be elected by a vote of the IM Committee members present at that year's annual IM Meeting (one vote per site). A majority vote of DataTask can call for a special vote of the entire committee at any time. The chairperson will serve as an ex officio, non-voting member of DataTask.

Susan Stafford was unanimously elected as chairperson. Susan's term will begin January 1, 1999. The IM Committee will request Network Office approval for the chairperson position and funding for the chairperson to attend the annual IM Meeting. *(Editor's Note: Approval and funding for this position was granted at the Fall, 1998 CC Meeting*.)

8.2 Election of New DataTask Members

A closed LTER Data Manager session was conducted to conduct DataTask elections. The vote was preceded by a discussion of whether DataTask should be expanded to seven members (from 6). It was noted that the creation of standing committees might relieve DataTask of some of its workload. The committee needs to find a balance between being large enough to have resources to do the work, and being small enough so that each member is responsible for responding. The committee decided the additional person was not needed.

Susan Stafford and Don Henshaw were rotating off three-year DataTask terms. Karen Baker and Don Henshaw were elected to new three-year terms. The current DataTask membership is as follows with year of election:

Susan Stafford (Chairperson, ex officio) - 1998
Karen Baker (PAL), Don Henshaw (AND) - 1998
Darrell Blodgett (BNZ), Michael Hartman (NWT) - 1997
Chris Wasser (SGS), John Porter (VCR) -1996
James Brunt (NET, ex officio)

8.3 Outside Visitor Policy for IM Meetings

A review of outside visitor participation at the annual LTER IM meeting was also conducted during the closed portion of the session. A proposal for outside participation was suggested as follows:

1. Add participants and invite guests as needed - consider funding for invited guests
2. Visitors should be limited to 20% of total or about 6 persons
3. Requests for additional participants should be sent to IM Chairperson
4. Encourage Bob Waide to attend
5. Encourage PI's to attend
6. Retain a closed session for at least some part of the meeting
7. Consider an open invitation symposium in the future

9.0 Feedback on the 1998 Data Managers Meeting (B.Benson (NTL))

Barbara Benson offered to collect and summarize feedback on the 1998 meeting from participants. Eight participants responded to the call for feedback. In general, the 1998 meeting was judged to be well organized and very informative. One participant commented, "It WAS a very productive meeting and I think we have set ourselves a healthy and productive agenda for the year ahead." Participants found the meeting energizing and a source of new ideas.

Participants had numerous constructive suggestions for improving the effectiveness of the data managers meeting and the operation of the group. Concerns were expressed about the balance between presentations and time for small working groups. An example offered was that the NACSE proposal discussion would have benefited from concurrent small working groups to generate collaborative proposals. Meeting time is at such a premium for the data managers, and one person felt an opportunity for more progress by standing working groups was missed

Ideas were offered to continue to broaden the level of involvement by all data managers in the meeting. This year more people were actually involved in presentations and writing. Quick round-table discussion encourages participation at meetings.

The importance of continuing site bytes was stressed (especially as a way to integrate new sites). It is a source of information on what sites are and are not doing. People liked having the site bytes available online prior to the meeting. It was suggested there still be presentation of site bytes at the meeting (2 minutes/site) focused on one question such as "what I'd like to hear about from others is xxx because at my site yyy".

The group needs to discuss how to handle the situation when a site does not send a representative to the data managers meeting. The suggestion offered was that lack of attendance be mentioned at the data managers and CC meetings.

The meeting web page needs to be set up well in advance. The page could contain a sign up page, logistics, information, site byte posting, and sharing of relevant reports and URL's. This organization is discussed further in the knowledge management section of this report.

The year 2000 meeting needs to be scheduled early to prepare for ESA (suggested place: SEV out of field season).

Suggestions made on the operating procedures of the data managers committee included desires for greater clarity about procedures and mechanisms to insure all views were heard. Clarification was sought in the allowable participation of multiple data managers from a site (voting in meetings and on data task). It was suggested that multiple members have fractional votes.

A desire was expressed for more extensive written guidelines for committee functioning (e.g., selection of working

group chair and length of tenure, procedure to create a new working group, procedure for working group members to dissent the chair's written report). " … Having brief written policies just makes things seem more professional and can provide a framework for better group relationships because the rules are spelled out."

Better way of chairing working groups and concerns about capturing all members views were raised as issues. It was suggested that working groups have two chairs: one gives the summary, the other takes notes and makes sure all opinions are presented.

We want to insure data managers attend and are on the agenda at CC meetings. One respondent wanted to a keep history of this involvement and thought it was important to have both the Executive Committee data manager and local site data manager speak at the CC meeting.

A new working group on network operations system was recommended to provide "input to and if necessary actual assistance to the network office in the operation of the network office computers … to make the network computers more useful to the LTER sites in very specific ways". This working group could handle such things as creating and maintaining mailing list functions for LTER sites that wanted them, handling web pages, web email archives, and shared code archives, and selecting and installing and/or evaluating of commercial software for use at LTER sites. The working group would be particularly useful to facilitate communication of software solutions between sites. For example, some sites might be willing to share their data logger filters, bibliography filters and various other utilities.

10.0 Meeting Plan for 1999

The 1999 IM Meeting is tentatively planned for August 5-7 in Spokane, Washington. The meeting will precede the annual ESA meeting scheduled for August 8-12, 1999 in Spokane's Opera House and Convention Center. A webpage for meeting planning has been established at http://www.lternet.edu/network/committees/data/

APPENDICES

A. List of Participants

(Note: For e-mail, use the usual LTERnet naming scheme, e.g. dhenshaw@LTERnet.edu.)

AND: Susan Stafford
AND: Don Henshaw
ARC: John Helfrich
BES: Christopher Steele
BES: Timothy Foresman
BNZ: Darrell Blodgett
CAP: Peter McCartney
CDR: NONE
CWT: Ron Rouhani
CWT: Ned Gardiner
HBR: John Campbell
HFR: Emery Boose
JRN: John Anderson

JRN: Ken Ramsey
KBS: Garrett Ponciroli
KNZ: John Briggs
LUQ: Eda Melendez-Colom
MCM: Denise Steigerwald
NET: James Brunt
NET: John Vande Castle
NET: Bob Waide
NTL: Barbara Benson
NTL: Robin Stubbs
NWT: Michael Hartman
PAL: Karen Baker
PIE: Hap Garritt
SEV: Gregg MacKeigan
SGS: Chris Wasser
VCR: John Porter

GUESTS:

Miguel Acevedo (IM Committee for the Latin America Network)
Rick Clutter (National Estuarine Research Reserve System (NERRS), Baruch Institute)
Chris Daniels (Baltimore Ecosystem Study, guest speaker)
Dick Olson (Oak Ridge National Laboratory, DAAC project)
Cheryl Solomon (NASA Global Change Master Directory (GCMD) project)
Local BES staff

B. 1998 LTER Site Flashes

- AND
- ARC
- BES
- BNZ
- CAP
- CDR
- CWT
- HBR
- HFR
- JRN
- KBS
- KNZ
- LUQ
- MCM
- NET
- NWT
- NTL
- PAL
- PIE
- SEV

- SGS
- VCR

Andrews Experimental Forest (AND): Don Henshaw, Gody Spycher, and Susan Stafford

The major emphasis of Andrews LTER information management efforts have been directed toward putting more Forest Science Data Bank (FSDB) data sets online. Following NSF directives for posting data online within two years with a minimum of restrictions, scientists and data managers have identified and evaluated current LTER proposal data sets as well as older legacy data sets as potential candidates for online status (Type 1 data). This year to date, eight new data sets have been added online as a result of this effort, and this effort is ongoing.

Concurrently with this effort, the Andrews Site Data Access Policy was rewritten and posted on the Andrews webpage (http://www.fsl.orst.edu/lter/datafr.htm). This policy requires users downloading data sets to abide by a Data Use Agreement and to fill out a User Registration Form. The Data Use Agreement asks users to 1) notify Andrews' site researchers when data sets are used in any derivative work, 2) acknowledge our support (an appropriate acknowledgement statement is included), 3) send copies of any resultant publications (to the address given), and 4) generally be good citizens with regard to the use of this data. Criteria have been developed that would allow specific restrictions to be placed on the access of certain data sets (Type 2 data). Conditions justifying Type 2 data status will be documented for these data sets, and we expect Type 2 status to be rare.

Using the Andrews Experimental Forest's 50th anniversary event, August 21st, as incentive, a bibliography of all Andrews-related publications over the past ten years will be printed and made available this summer. In preparation for this effort, a hierarchical keyword list was developed and all publications were keyworded in a joint effort of researchers and data managers. The keyword list is part of the ongoing effort to support online searches of publications and spatial and non-spatial data objects by author, keyword, species name, and geographic location. Databases will reside in SQLServer (Pentium II with 12 Gb RAID array) and web searches will be directed through the Internet Information Server (IIS) webserver. The data model for this project has been constructed and will be maintained in Logic Works' ERwin database design tool. Don Henshaw, Gody Spycher, and Hazel Hammond have been consulting with UDP Solutions Center software engineers on the data model and design for our metadata.

A NSF-funded Supplemental grant will be used to help purchase a new webserver for the LTER and OSU's Forest Science Dept. A Sun Ultra 10 with 90 Gb will run Netscape Enterprise Server. Additionally, this grant will also fund a student to assist with putting Andrews databases onto the Web.

In October 1997, Don Henshaw attended a CLIMSTAN workshop funded through the Network Office that brought together climatologists, data managers, and data users for the purpose of revising the LTER meteorological observation standards and to formulate standards and procedures for making LTER climate data publicly available. Henshaw continues to play an active role in the development of ClimDB, which is a prototype method for sharing distributed data sets over the Web.

In February 1998, Susan Stafford, and Henshaw hosted a collaborative meeting of LTER DataTask Information Managers with the Northwest Alliance for Computational Science & Engineering (NACSE). Further collaboration is likely to follow with LTER IM's using NACSE web and database tools to enhance sharing distributed datasets. Additionally, Cherri Pancake (OSU Computer Science Dept.) and Stafford have been funded by the Database Activities program at NSF for three years ($872K) to specifically address the data integration needs of independent, small-scale databases maintained by geographically disperse research teams. There are 17 interdisciplinary databases forming the Oregon Coalition for Interdisciplinary Databases (OCID) which serves as the testbed for this project - demonstrating how the software can be deployed to cross disciplinary, agency, and hardware/software boundaries to create a virtual database of biological, geophysical, and climatic data. The data integration needs of the LTER network will be addressed by this project, specifically using enhanced versions of HyperSQL and QueryDesigner software

packages developed with this project.

Stafford is also an editor of Data and Information Management in the Ecological Sciences: A Resource Guide with William Michener and John Porter. The Resource Guide was a product of a two-day workshop jointly funded by OBFS, LTER, and NSF (August 8-9, 1997) entitled "Data and Information Management in the Ecological Sciences" that was held at the University of New Mexico in Albuquerque. Andrews' personnel Stafford, Henshaw, Hazel Hammond, and Fred Bierlmaier authored or co-authored three papers in this guide.

Arctic Toolik Lake (ARC): Jim Laundre

The ARC LTER project was renewed this year for another 6 years. Following is an excerpt from the proposal on the Research Approaches of the Arctic LTER Project: The goal of understanding the present and predicting the future of a complete ecosystem is an ambitious one that can only be approached with a long-term plan of research. The elements of the plan include: 1. long-term monitoring and surveys of natural variation of ecosystem characteristics in space and time. 2. experimental manipulation of ecosystems which are maintained and measured for decades 3. synthesis of results through modeling at ecosystem and watershed scales 4. feedback from the synthesis and modeling to identify gaps in knowledge.

Towards these goals we continue to add experimental and monitoring data to our on-line database. Various models like MBL-GEM, TEM, TOPMODEL, plus others are making use of the ARC data to scale up to larger watersheds and ecosystems across the North Slope of Alaska.

Donald (Skip) Walker, INSTAAR, University of Colorado now has ARC/INFO coverages of ARC LTER and NSF/ARCSS/LAII-funded study sites available online (see http://www.colorado.edu/INSTAAR/TEAML/atlas/). Maps of Alaska North Slope study site (1:250,000 scale) and Toolik Lake study site (1:25,000, 1:5000, 1:500 scale) are available. Data layers include: vegetation, geomorphology, hydrology, elevation, soil carbon, percent water, and geology.

A new project is The LTER Alaska Transect project, which is a joint venture of the Arctic and Bonanza Creek LTER projects. The overall goal is to forge links between these two projects through a joint effort to collect spatially-distributed data along a transect cutting across the entire state. Both aquatic and terrestrial measurements will be made with maximum use of remote sensing (AVHRR, SAR, LANDSAT) and continuous recorders (weather stations, stream flow and temperature). Responses of species and processes to different conditions will enable us to compare ecosystems that experience summer temperatures similar to those predicted for the Arctic within the next century. The data will also be used to test model predictions and expand our understanding of the spatial variability of northern Alaska.

Through money from a LTER supplement proposal we will be replacing our UNIX server with a NT server. This system will facilitate building relational databases with web front ends, allow easier sharing of data among investigators, allow easier sharing of GIS data, and allow integration with other GIS users at MBL who are running NT ARC/INFO.

John Helfrich will again be representing the ARC site at the annual LTER IM meeting. Jim Laundre could not attend. Late July and beginning August is peak field season for the ARC LTER site.

Baltimore Ecosystem Study (BES): Chris Steele and Tim Foresman

We are closing in on the first anniversary of the Baltimore Ecosystem Study (BES) http://baltimore.umbc.edu/lter. Our first field season is under way and things are looking good. The renovations of the BES Offices are just about

complete, and we are set to move in during or just after the 1998 ESA meeting. We have enjoyed a tremendous amount of support and cooperation from PI's and Collaborators of the BES. The information and data management staff of the BES has been quite busy setting up internal computing and communications networks, information and data management protocols, and task management protocols. We have purchased a high-end Dell Workstation 400 (dual 333 processors, 256 MB RAM, 18GB HD) to act as our server. This is a very fast/powerful machine that will serve us well for the next 3-4 years. We have also purchased various peripherals to compliment the computing infrastructure of the BES.

We have established a series of working group listserve groups that facilitate communication between members of the study interested in particular topics. In terms of web development and updating, we are taking an "administrator friendly" approach. In cooperation with Cybergroup (a firm based at the University of Maryland, Baltimore County), we are developing an interface that will allow quick and easy updating to the BES web page. An example is the PI who will be able to update the BES calendar from his/her local PC. The update is sent to a pending area where some member of the data management staff will confirm that the update is valid by a single mouse click. This virtually removes the data management staff from routine updates that PI's can do with very little training. Updates such as calendars, bibliographies, personnel information, and a site glossary will be designed with this approach.

Additionally, in cooperation with the Baltimore Washington Regional Collaboratory (http://www.umbc.edu/bwrdc) (NASA funded) we will be adding an ecological component to version two of the three-tiered General Entry Metadata (GEM!) tool. We will extend the tool to include data sets uploading and downloading functionality. The BES data management staff has been spending a large amount of time and effort on developing an internal operations page where PI's can retrieve data, meeting notes, sub group updates, and general information that is not meant for public sharing. Our strategy is to equip BES personnel with the data, information, and tools that will allow them the ability to "do science" without having to be computer scientists or IT professionals, while ensuring that the data management staff is not overwhelmed with support tasks for the entire site. The BES data management staff is also working to link our data to the Human Ecological System (Machlis, Force, and Burch), which is the synthesizing model for the work of the entire Baltimore Ecosystem Study. Since this is such a novel approach, we are proceeding cautiously so as to meet the desired goals of the Human Ecological System as a synthesizing model as well as a data serving interface. We will dedicate a large amount of time to this effort in the very near future.

Bonanza Creek (BNZ): Darrell Blodgett

Our renewal proposal was sent in in February and received some mixed reviews. We are now in a probationary period, and will have to resubmit a new proposal in eighteen months time.

Terry Chapin our LTER Project PI has now moved back to Fairbanks bringing with him an entourage of postdocs and grad students.

A New Co-PI Forest Service Unit Leader :Marylyn Walker a plant community ecologist from the University of Colorado will be co-director of the LTER program. She plans to move to Fairbanks next spring/early summer, but will make several trips to Fairbanks over the winter

Frostfire Project: A joint project between Bonanza Creek Long Term Ecological Research Program (UAF), Pacific Northwest Research Station (U.S. Forest Service) and the Alaska Fire Service (Bureau of Land Management). A very large prescribed burn project involving many people. Personnel information and mailing list's were created and are maintained on Bonanza Creek's web site. Other Frostfire information is split between the Pacific Northwest Research Station's web site: http://www.fsl.orst.edu/home/usfs/gepp/alaska/frstfire.htm and the Bonanza Creek web site: http://www.lter.alaska.edu/~jirons/cpcrw_www/frostfire/frostfire.htm

The Frostfire prescribed burn has now been put off until next summer due to wet conditions in the burn area.

Darrell Blodgett Attended the CLIMSTAN Workshop in October 1997, and the February meeting between LTER Data

Task and The Northwest Alliance for Computational Science & Engineering (NACSE). Keep those winter meetings coming! Keep summer meetings to a minimum. The salmon are running!

During the CLIMSTAN Workshop several Data Managers in attendance: Karen Baker, Barbara Benson, Darrell Blodgett and Don Henshaw discussed the idea for SiteDB. SiteDB is an attempt to create uniform views of all the LTER sites containing key parameters for that particular view (Site, Climate, Soils, Vegetation...). Individual sites would retain control of the information that is displayed through web forms used to update the SiteDB database. This database would also be searchable via a web interface so users could pull up selected parameters for all sites or for a subset of sites.

A SiteDB prototype was developed at Bonanza Creek with the help of Karen Baker, and Don Henshaw. Later, other members of Data Task also participated in the testing of the prototype. Prototype's URL: http://www.lter.alaska.edu/cgi-bin/w3-msql/sitedb/site_menu.html

Sometime this past spring a prototype for the ClimDB metadata database was also created at Bonanza Creek. The prototype used the structure formulated at the CLIMSTAN meeting in October 1997. Prototype's URL: http://www.lter.alaska.edu/cgi-bin/w3-msql/climdb/climdb_menu.html

We are currently using MiniSQL behind Netscape's Enterprise server to store and display our site bibliography, personnel directory, project and data catalog, and information on individual research plots/stations. This information is also maintained and updated via the web.

Cedar Creek (CDR):

<None>

Central Arizona / Phoenix (CAP): Peter McCartney

The past year has been a learning experience at CAP LTER, and no less so for the data management team. Our primary goals for the first year have been:

- Establish an information management infrastructure that will support the diverse and widely distributed research and outreach activities of CAP LTER,
- Consolidate existing GIS, bibliographic and research datasets on campus to provide a foundation for new research in the Phoenix urban environment,
- Work closely with project leaders of the pilot studies to develop a set of procedures that fosters early and continued involvement of the data management team in database design and research activities.

Resources:

Computing activities at CAP LTER are carried out over a distributed network of resources. At the core is the computing facility maintained jointly by the CAP LTER and the Archaeological Research Institute (a collections and data repository within the Department of Anthropology). In January, Ted Oliver was hired as staff to assist McCartney in database management and system administration. The lab currently has approximately 1500 square feet devoted to computing activities. The LTER server is a PII 300mhz equipped with redundant power, cooling and storage (RAID 5). Current capacity is 27 gigabytes. Backups are made to tape following a 3 week rotation. This server is used for the LTER data archive, web and FTP services and general user file services. In addition to personal computers used by the data manager and staff, the lab offers 6 Windows NT workstations for research and several Windows 95 machines used for data entry. We also received two Macintosh clones that were part of a larger corporate gift from Motorola.

The database server for CAP LTER is Microsoft SQL Server. Applications programming is done using Microsoft Access and Visual Basic. Databases are accessed through the web service using server-side scripting (MS Active Server Pages). GIS software includes ESRI ArcView, MapObjects, and ERDAS Imagine. X-windows software is provided for running remote sessions on the university Unix systems.

The Information Technology GIS lab is a partnering facility directed by Jana Fry (co-leader of the Data/GIS/RS core team). It provides Unix and NT access to ArcInfo, ArcView, and Imagine. The lab also provides high-quality input and output devices. Because of pre-existing working relationships with many LTER community partners, the GIS lab is a primary channel for data-sharing between these sources.

Finally, the Geological Remote Sensing Lab directed by Phil Christensen receives and processes all remote sensing data used by CAP LTER. This lab is a mixed Unix and NT environment using ER Mapper and ERDAS Imagine as the primary software tools. This lab is currently developing a remote sensing server which may be used as the primary distribution medium for LTER image data.

Procedures:

A general procedure for designing, implementing and documenting project database was developed and has been applied to six pilot projects. After consulting with the PIs, a data modeling tool (ER/Studio from Embarcadero) was used to produce a generalized schema of the database complete with entity and attribute metadata. One advantage to this type of software is the ability to deploy a database schema on several physical platforms. For CAP LTER, most databases are built in MS Access, then migrated to SQL Server after design becomes stable. Interfaces are developed with Access which features very rapid development time, rich options for quality control, and easy redirection of forms when tables are migrated from Access to SQL Server. Completed databases and entry applications are delivered in less than a week after the initial meeting with PI. After initial development, draft metadata in the form of data dictionaries are posted on the data management section of the intranet site.

Data management has also been involved in projects that seek to acquire and import existing data from external sources. Research Assistants employed by such projects frequently work with these datasets at the LTER lab so that we may provide support for these activities.

Considerable effort has gone into designing a comprehensive database to hold metadata for CAP LTER datasets. Three sources were considered: Michener et al. paper on metadata, the Federal Geographic Data Committee, and an informal survey of existing metadata documents on other LTER sites. Our strategy has been to use the Michener paper as a conceptual guide to the different levels of information required, while adapting the specificity of the FGDC standard wherever applicable (e.g. handling keywords, projection information, entity and attribute descriptions). Once a draft format was developed, it was compared to existing LTER metadata to ensure maximum completeness. The draft structure of the database has been posted in the form of an Entity-Relationship diagram and a data dictionary.

Based on the LTER data managers' draft, a data access policy was developed for CAP and is awaiting final approval. It will be published on the data access page of the web site, along with a mandatory login form for visitors wishing to download data.

Products:

Two immediate tasks were to define the core databases needed to manage LTER data and to set up the CAP LTER website to provide a medium for accessing information. A series of integrated databases were defined and implemented as web applications: a bibliography based on a previously compiled list of environmental literature for central Arizona, a projects listing personnel directory, and a current events calendar. The dataset catalog and associated web application is nearing completion and should be online within a few months. This application will rely on database calls for both the data tables and the metadata. The plan is to design metadata output to follow several formats on demand including the LTER exchange format introduced by Porter last summer, FGDC, and any future LTER standard. Because the participants in CAP LTER are distributed, the web site has played in important role in local, as well as external, communication. An Intranet site was created for posting internal announcements, meeting minutes, and extensive computer support pages. A series of NNTP discussion groups were created and participants

have been encouraged to use them so that a record can be made of our electronic dialogs.

Two projects, one directed by McCartney and Fry, the other by Phil Christensen and Mike Ramsey compiled existing GIS framework data and remote sensing imagery to provide a spatial framework for initiating LTER research. Available datasets have been posted for web and ftp access; in addition, websites were created to summarize and preview GIS and remote sensing data.

LTER data management was active in external grant efforts. McCartney was coPI on three proposals that involve collaboration between the ARI/LTER lab and various departments. One, a proposal to NSF for vBNS connectivity was funded in November. Two other NSF proposals are pending: one to the Archaeology program to prepare metadata and internet publication of a 30+ year research project on the archaeological site of Teotihuacan, Mexico, and the other to the URI program to study social factors associated with relative success of several water-based community development projects. McCartney and Fry also co-authored the data management portion of a proposal to the EPA Empact program to bring environmental information to the public.

McCartney offered a course in GIS in Archaeology in the Spring semester, which was attended by several students from the LTER program. In response to requests from PI's, a modified version of this course focussing on GIS and Database management for ecological research may be offered in the following spring.

Directions:

Top priority for next year is getting the dataset catalog online and to publish the datasets that are ready to go online. We will also be adding a web-based GIS application developed with Visual Basic and MapObjects to provide more interactive display of GIS layers from the study area.

In response to PI feedback, a new feature to be added in the fall will be a regular series of training workshops designed to familiarize LTER scientists with the resources and data available at the LTER lab. These will include primers on accessing the LTER, basic SQL database query, and using simple GIS tools to produce data-rich maps. There will also be training on using data entry applications developed in MS Access.

With a supplemental grant for curation from the LTER program, we will expand the current catalog databases to include inventories for specimen collections and for reference collections from the Biology department. These will parallel similar curation catalogs currently maintained for the archaeological collections.

Finally, new solutions for getting more complete and rapid information on project research designs and methods will be sought. Both web-based forms for entering project description and more aggressive follow-up with PIs will be considered. The existing project descriptions which are database-derived, will be augmented with static web pages for projects that provide more visual and narrative information.

Coweeta (CWT): Ron Rouhani

The Coweeta Site has gone thru some major changes. I have replaced Gil Calabria as the new data manager. During the past 9 months the following changes and additions have been performed. Coweeta's Webpage has been completely changed. All Datasets have been moved to a MSQL Database, which allows the users to query the datasets online. A total of 9 new datasets have been added to the "Ongoing Research" page. Monthly updates to the ongoing datasets are performed by perl scripts. The 1030 entry Bibliography was placed online. The users may search thru the bibliographies by using a specific or general query string. A fully interactive web page of our plant tissue and soil sample archive has been placed online. A metadata page was created so that all protected dataset metadata would be available to the public. The page allows the researchers to display their ongoing work. An electronic version of the Coweeta brochure was placed online. A 12 Gig. Disk box was purchased and installed on the server. A needed power supply was purchased for the server and its disks.

Harvard Forest (HFR): Emery R. Boose

Recent developments at Harvard Forest include installation of three new Windows NT workstations (Dell Optiplex, 128 mb / 12 gb), and near completion of an online, searchable bibliographic database of Harvard Forest publications (nearly 1000) using MiniSQL. A prototype of an online GIS atlas supporting graphics display and query has been developed this summer as an REU student project, and work is beginning on creation of a data catalog for the new Sample Archives. During the coming year we plan to upgrade the speed of our network connection to the University, install one or more Windows NT servers, and (if all goes well) move the Harvard Forest web page from LTERNET to a local NT server.

Hubbard Brook (HBR): John Campbell

Hubbard Brook data are now accessible at http://www.hbrook.sr.unh.edu/. Shortly after this web page was made public, the Hubbard Brook gopher site (hbrook.unh.edu) was discontinued due to the limited storage capacity of the computer and limited space for the computer at the University of New Hampshire (UNH). The web site and data are now maintained on a UNIX-based server at UNH. Originally, the web page was to be transferred to a new IBM server that was installed at the Forest Service Laboratory in Durham. However, a firewall that was installed for security, made this impossible. In January 1998, we obtained permission to let the web site remain at UNH where it is maintained remotely by John Campbell at the Forest Service. The following Hubbard Brook cooperators have posted data and information on other servers that are linked to the main Hubbard Brook web page: Institute of Ecosystem Studies (IES), Syracuse University, USGS, and Yale University.

The main Hubbard Brook web site currently consists of 38 data sets, which is a substantial improvement over the 10 sets that were posted last year. The web site also contains data from a greater number of watersheds compared to last year and includes more recent years. Long-term routine hydrometeorologic and other Forest Service data are now current through 1997 and are posted annually after QA/QC analyses are performed. We are now in the process of modifying our data access guidelines which should be completed shortly.

In 1997, a top priority was to update the physical sample archive and to barcode and catalog samples that had been accumulating since 1994. In order to facilitate this process, the existing database was transferred from DBASE IV (DOS) to MS Access (Windows). We have been working toward eliminating the backlog and have recorded and archived 1200 new samples this year including soil samples, water samples and our first lysimeter samples. The remaining samples will be archived by the end of the summer at which point we will be able to add new samples as they are submitted. Within the coming year we will continue to develop the new software so that the process of searching and retrieving sample data is more automated. Currently, a portion of the physical archive database can be searched on the web and in the future we would like to include the entire database.

During the past year a Hubbard Brook GIS (Arc/Info) database was developed at the Forest Service in Durham and is stored on a UNIX-based computer. This database consists of over 20 coverages with corresponding metadata from many sources including Cornell, USGS, and UNH. These data have recently been updated and the metadata have been standardized. All coverages are available to the public and can be viewed and downloaded from the Hubbard Brook web page in arc export format.

Jornada (JRN): Ken Ramsey, Barbara Nolen, and John Anderson

Ken Ramsey has been busy upgrading both hardware and software for our file server and client workstations. Our file

server operating system has been upgraded from Netware 4.10 to IntraNetware 4.11. This upgrade allowed adding UNIX services such as FTP and TCP/IP. It also allowed us to switch from Glaci HTTPD web server to Netware Web Server, which enhanced operation, security, and manageability of our web server. Support has also been added for MacIntosh client log-in, print, and file services for visiting researcher's MACs.

Our Mercury Email server software has been upgraded to ensure that our email server can not be used to regenerate, or autoforward, SPAM. Several add-on options have been added to ARCserve backup software as well as upgrading ARCserve. These allow it to work with IntraNetware as well as provide enhanced performance. The main options added to ARCserve were the Disaster Recover Option and the Windows NT Client Option. These options provide for rapid file server operating and file systems recovery in the event of hardware failure as well as adding NT Workstations to the automated backup routine. We are still using ARCserve software and Exabyte EXB-10h 8mm Tape 'Jukebox' to perform fully automated nightly backups of file server, LAN clients, and Lab PCs.

In an effort to minimize the effects entering the new millenium we have conducted an inventory of our PCs, file server, and laptops. We have also initiated a hardware and software inventory. This inventory has two goals: one is to add a hardware component and software database to our Information Management System (IMS) and the other is to identify which software packages are not Year2000 compliant.

Ken Ramsey is investigating alternative ways to upgrade our current and future server operating systems and components. There are several existing state and education contracts that we can 'piggyback' onto that can lower overall cost. By reducing costs of server operating system upgrades and spreading the costs out over two-year site licenses we can lower the Total Cost of Ownership(TCO) while adding greater flexibility to our web site in the years to come. The flexibility is derived by the opportunity to add to the various add-on components that we would have available to us without any added cost normally involved with 'shrink-wrapped' components using traditional purchasing methods. These licensing agreements would be with Novell and Microsoft and will also give us multiple options for approaching any solution into the foreseeable future. This approach of being able to budget upgrades on a fiscal basis will allow for future enhancements to our LAN while minimizing worry about additional costs, as the costs associated with these options are already accounted for. Ken Ramsey has also begun to research vBNS and how our site could benefit from the increased access speed and security that vBNS could offer to our site. Any information that other Information Managers within the LTER network can share on how they have, or are going to obtain, NSF funding jointly with their educational institution ties would be greatly appreciated.

We have requested supplemental funding from NSF. A portion will be used for a new web server, web application development software, and web database applications to allow data query-ability of data and metadata. Ken has also researched various combinations of database software and programming languages that will be used for developing both our upcoming web application and to enhance our current IMS. We hope to increase the use of relational databases to perform routine IMS functions as well as enhance our QA/QC capabilities of current and future datasets. One of the first databases we hope to start with is metadata for our site's growing GIS and Image archives which will comply to the FGDC metadata standards.

We continue to add new data sets and update ongoing ones on our web page.

As we implement the hardware and software changes that will allow the development of our web and LAN applications, we will also be changing the style and adding to the content of our web site. Another major benefit of the supplemental funding is that we will be able to hire a fixed-term part-time employee to help in getting our GIS/Spatial metadata online to facilitate access and query of our GIS and image archives through our web site.

We have added a new PII-300 computer for use as a GIS/Spatial PC. Upgrades have been made to all GIS/Spatial software packages to enhance the performance and capabilities of software used by our Spatial Analyst, Barbara Nolen. Barbara has been the leading force in establishing New Mexico State University (NMSU) site licenses for ESRI ARCinfo and ERDAS Imagine software packages. This benefits both our site as well as other departments of NMSU by increasing software capabilities and options while significantly lowering costs. One of the options we are considering implementing is ESRI Image Map Server, which could provide web access to our map images and their associated metadata. This would reside on a separate computer from the web server because of intensive CPU usage.

Barbara Nolen received the 1997 USDA ARS Natural Resources Institute's Collaborator of the Year Award for her outstanding research collaboration efforts in the JORNEX and PROVE campaigns on the Jornada Basin. Congratulations!

Barbara continues efforts to build metadata for the gis and remote sensing database. We are also continuing to collect AVIRIS data over the Jornada and Sevilleta. GIS, GPS and Remote Sensing overviews have been shared through the K-12 Educational Outreach program through the USDA. Color-Infrared photography is being utilized for a current vegetation map covering the USDA Jornada Experimental Range (JER) and the Chihuahuan Desert Rangeland Research Center (CDRRC). Landsat Tm images are being used to continue the current soils map of the Desert Soils Project to cover the USDA and CDRRC. This mapping project will also produce a geomorphology map of the basin. Barbara is also involved in rescuing field notes for vegetation map from 1928-1929 over the JER. She continues interagency cooperation on all fronts.

Kellogg Biological Station (KBS): Garrett Ponciroli and G. P. Robertson

Data management highlights at KBS include the move from a VAX based system to an NT server, additional web-based data management capabilities, a new GIS resources library, and the development of an on-line herbarium. Our move to an NT server is almost complete, and our URL has changed to http://lter.kbs.msu.edu. Web-based data management improvements include a gradual move from flat files to searchable database queries for our more complex data sets (NPP, trace gases, weather). Most of the weather data has already been moved to an Access97 database queriable from the web (http://lter.kbs.msu.edu/weather). We have also put on line a GIS resources library for KBS (http://lter.kbs.msu.edu/GIS), and have recently received supplemental funding to place our herbarium and insect collections on-line with museum-quality images downloadable from a web-based catalog.

Konza Prairie (KNZ): John M. Briggs

As with most other LTER sites, this past year we have increased our efforts to get ALL LTER data on-line within the two-year window. This has been the KNZ data management policy, since the last renewal, but this past year we have made a special effort to get even more data on-line, with the adoption of the new LTER data policy. That policy (http://climate.konza.ksu.edu/general/temp/lterpolicy.html) states that all LTER data will be on-line within 2 years after collection. In order to help us in this endeavor, we hired a 1/2 research assistant. This has been successful and over 90% of the Konza LTER data is on-line and freely accessible over the WWW. Plans are for now to start getting all of our spatial data on-line including all of the metadata associated with it.

Our server (http://climate.konza.ksu.edu) continues to be an outlet for the data files of the Konza Prairie LTER program. We are receiving over 100 non-KSU visits to our site. Since it is a major part of our LTER effort, we are upgrading the Internet line to our WWW server to a full 100-mps line. The Konza Prairie data management staff continues to be involved in network wide activities. One major item included developing the new LTER data access policy that was voted in by the LTER CC last October. In addition, Briggs presented two talks at the symposium Data and Information Management in the Ecological Sciences that was held in Albuquerque, NM prior to ESA last year. That symposium resulted in an electronic book entitled: Data and Information Management in the Ecological Sciences: A Resource Guide. W. K. Michener, J. H. Porter, and S. G. Stafford (eds)

(http://www.lternet.edu/ecoinformatics/guide/frame.htm), which should be a valuable guide to the Ecological community.

We also requested funds to purchase a new server to handle our research program. At this time, we have not made out final decision whether to go with NT or Novell 5.0 as we are still reviewing both software and hardware options. Other hardware/software news includes that our field site, Konza Prairie, should have a full T1 connection by the end

of this month.

Luquillo (LUQ): Eda C. Meléndez-Colom and John Thomlinson

REMOTE SENSING TECHNICAL BYTES

We now have a catchy name for the remote sensing/GIS lab: the Spatial Analysis Technology Lab at the Institute for Tropical Ecosystem Studies, or SATLITES. Unfortunately, the acronym does not translate so well into Spanish (LTAEIEET).

In the last year we have installed a new SPARC Ultra-2 workstation, and we are currently in the process of porting the application software to it. The workstation contains two 300 Mhz processors and a 33 GB RAID disk array, which we estimate should take several months to fill to capacity. We will continue to use our old workhorse, sunceer, primarily for email and web applications. There are currently three dedicated Xterms (one 17", two 19"), and four PCs (three 17" and one 15") with Exceed emulation software used to access the UNIX system. We have purchase orders out for three 21" PCs for image processing/workstation access. We are attempting to move away from Imagine (TM) software for our image processing and have just ordered UNIX and Windows licenses to ENVI (TM) as a first step in that process. Our other notable hardware upgrade has been the addition of an HP 755CM plotter to replace the old Calcomp pen plotter.

In terms of personnel, we have hired a systems manager for the lab and we are in the process of adding a full-time GIS technician. In addition, we have had one undergraduate and two graduate students for the past year, and we have added two more undergraduates for the summer. The number of students with senior theses completed in the SATLITES lab is now five, and we have two masters theses completed and three underway, in addition to a major portion of a Ph.D. dissertation. The lab has also provided data and analysis assistance to one graduate student in the U.S. who recently completed her thesis.

We are building a reference library of ground control points for Puerto Rico, in cooperation with the International Institute for Tropical Forestry (IITF), starting in the area of the Luquillo Mountains. The recent acquisition of a Trimble Pro-XR has made this possible. We are also working with IITF and off-island LTER researchers to locate study sites as accurately as possible. We have prepared maps of the El Verde research area to help investigators locate new study sites and understand treatment histories at existing sites. We continue to support the LTER program by helping with spatial aspects of the ongoing Hurricane Recovery Plot study, plotting maps of the El Verde area and LEF for proposals and manuscripts, and providing information on land-use patterns in the greater LEF area. We continue to develop and acquire island-wide data sets, although acquisition of high- quality satellite imagery has been a frustrating exercise. We are hoping to benefit from improved access to imagery from EOS-AM and Landsat-7 satellites through our close cooperation with the UPR Mayaguez campus and their new image downlink system.

DATA MANAGEMENT BYTES

After our site review last June, this has been a year of revision and enhancement of our existing data management system. A committee of three was formed, among our on- island investigators, to revise the documentation of our data sets. Letters are still being sent to the investigators asking for the update and/or completion of their data sets' metadata. Two thirds of the catalogued data sets have been revised as of June, 1998. In an effort to ease and stimulate the completion of the data sets' documentation forms, I met for three days with John Porter, previous to Luquillo's LTER Annual meeting, to learn how to program on the Web site using the Lite scripting language and the miniSQL query language. The purpose is to automatize the process of database cataloguing, to automatically produce an on-line version of the completed data sets' documentation forms and to update an on-line version of our data sets catalogues. The three levels of documentation forms were developed for this training and are already accessible from the Web site, but the programming part will be completed later this year.

The addition of new data sets to our catalogued has been slow due to the fact that we are concentrating more in the

revision of our existing documentation and enhancement of our system as described before. Nevertheless we have added two new data sets to our catalog. From the period of June 1 1997 to May 31, 1998 we answered a total of 17 data sets requests. (Nine of the data sets retrievals requests were from LUQ 4 investigators and collaborators, 5 data sets retrievals requests were originated by 3 LUQ students, and 3 data sets requests from 5 external investigators and students. Two of the local requests involved data manipulation). During the same period we have answered to 6 special requests, including 2 from the LTER Network Office.

Since last June 1998, I took part in two workshops for the Latin American - ILTER which involved the design and summary of a survey on the existing computer and telecommunication facilities of the potential and existing research stations of the different Latin American countries. The web pages at http://cesimo.ing.ula.ve/lalter and http://www.oikos.unam.mx/lalter (both mirroring each other) contains the preliminary results of this survey as well as the report of the different working groups of the meeting held at Venezuela from June 1 to 6, 1998.

Remote sensing bytes prepared by John Thomlinson Data Management bytes prepared by Eda C. Meléndez-Colom

McMurdo Dry Valley (MCM): Denise Steigerwald

The McMurdo LTER web site has been extensively expanded during the past year. In late summer, 1997, the web server was set up at the Institute of Arctic and Alpine Research (INSTAAR) in Boulder. Data and images were transferred from the Desert Research Institute (DRI) in Reno, Nevada to INSTAAR. However, multiple data sets not yet in the system at DRI needed to be processed in order to be compliant with the National Science Foundation (NSF) standards of providing data to the public after 2 years. The number of on-line datasets has expanded from 15 files available to the public in July, 1996 to 105 in July 1997, and 46 files in the restricted portion in July, 1996 to 120 in July, 1997. Datasets are provided on-line as ascii files, linked to the metadata pages using HTML coding, and organized by research topic. Topics currently include meteorology, stream location/measurement sites, stream hydrology, stream biology, experimental stream data, stream chemistry, limnological biology, limnological chemistry, lake ice, glacier locations/measurement sites, glacier chemistry, glacier mass balance, glacial meteorology, soil experiments, and bathymetry.

Once research datasets were on-line at INSTAAR, development of the web site expanded to include sections for maps, photos, publications, conferences, and personnel. HTML coding was used for most of this. Tools developed at DRI for a searchable bibliography, map-maker, and meteorological data extraction are still used on the web site at INSTAAR. A hydrological data extraction tool developed at the Wisconsin USGS office is also linked to INSTAAR's web site.

In the spring of this year, INSTAAR received some supplemental funding from the NSF to allow for purchase of more up-to-date software and hardware. We recently invested in a Dell OptiPlex GXA with Windows NT and lots of space for storage of data. This will be used in conjunction with our UNIX Sun / Ultra-Enterprise 150 Server. Software packages include Microsoft Office 97 and Visual Foxpro, making it easier to read files submitted to INSTAAR. In addition, we have acquired a zip drive, flatbed scanner and a slide scanner. This should speed up processing and allow more images to be used on the website. The supplemental funding also allows investment in Oracle software. Transfer of the database to Oracle is now underway. Denise is working on the design of the system for Oracle, which involves going through all the MCM LTER data sets, mapping entities to tables, attributes to columns, unique identifiers to a primary key, and relationships to foreign keys. Through the process, she is making sure that the existing data sets actually match the constraints and specifications that are being layed out for the Oracle tables. Many of the data sets are being "cleaned up" through this process.

Once the datasets are transferred to Oracle, tools will be developed to allow the web-user to retrieve data from various tables and combine them in one output file. Right now, (with the exception of the meteorological data extraction tool, and the hydrological data extraction tool and their corresponding files), if a user desires such a file, they must request it from the data manager. Development of data extraction tools will make the MCM LTER web site more user-friendly.

Transfer of the MCM LTER data management setup from DRI to INSTAAR was done in an effective manner. This helped contribute to the grant renewal that occurred in April. High praise was given for all aspects of the research, especially for the organization, thoroughness, and excellent response to previous criticisms and suggestions. Plans include building upon and improving things that have been set up, as well as developing tools that are desirable but not yet in place. Effective means of extracting information from datasets, maps, and miscellaneous information that relates to the research occurring in the McMurdo Dry Valleys is an ongoing goal for the data management of MCM LTER research project.

Niwot Ridge (NWT): Michael Hartman

After submission of our renewal proposal and subsequent addendum, the NWT LTER has been renewed for another 6 years. The full text and graphics of the proposal are available online at http://culter.colorado.edu:1030/Niwot/98renewal/index.html.

In addition to our continued work in the alpine, this next phase of our program will expand into the subalpine. This will not only provide new ecosystem comparisons, but also generate a separate line of questions regarding the biotic and biogeochemical significance of ecotones. As part of this expansion, a subalpine subnivean shelter was constructed this summer and is already providing preliminary hydrological and climate data. This shelter is a larger replicate of the shelter built on Niwot Ridge in 1993, and will provide a focal point for subalpine snow hydrological research. Other improvements made on Niwot Ridge this summer include the pulling of fiber optic cables to the Tundra Lab and the alpine subnivean shelter. This will enable us to provide current meteorological and hydrological data online from the Saddle, our primary experimental research area at 3525 m. Roof improvements were made to the main laboratory building at the Mountain Research Station (our field headquarters), and construction was begun on the new hostel.

The NWT LTER WWW continued to expand this year, with the addition of more available data sets, as well as increased intra-site use. We utilized protected areas of our server to enable investigators to view updates, post comments, etc., so that we were able to put together our renewal proposal online. We also developed "virtual field trips" for K-12 students and the general public so they can visit the alpine without enduring the thin air and high winds (http://culter.colorado.edu:1030/Field_trip/). Our data area (http://culter.colorado.edu:1030/Niwot/Niwot_Ridge_LTER_data.html) was revamped this year, including the addition of a data registration form. Our data management policy was also revised in order to better meet the NSF directive for making data available within 2 years (http://culter.colorado.edu:1030/Niwot/Niwot_Ridge_LTER_datmanpolicy.html).

The data management program received excellent marks in our renewal reviews, and we hope to continue to improve and expand. The Institute of Arctic and Alpine Research has finally conceded the need for institute wide computer administration support and will be hiring 2 full time people. This should greatly improve our productivity and free up a lot of time that is being wasted wrestling with system problems (many of which are not data management or even LTER related). We have begun collaboration with the Alexandria Digital Library Project on our recently acquired DEM and orthophotos of the Niwot Ridge and Green Lakes Valley area. Also, this summer a high resolution GPS unit (+/- 70 cm accuracy) was purchased, and will be used to accurately map the locations of all known plots (past and present).

North Temperate Lakes (NTL): Barbara Benson and Robin Stubbs

Database development:

Primary goals of the NTL-LTER information system are to (1) maintain database integrity, (2) create a powerful and accessible environment for the retrieval of information, and (3) facilitate linkages among diverse data sets. To meet these goals, we have used the Oracle relational database. At this time, most of the core data reside in Oracle. During

the past year we have incorporated most of the LTER zooplankton count data into the Oracle database. This data set is large and complex. Table design accommodates metadata and multiple levels of taxonomic resolution. Additional data sets have also been loaded into Oracle including historic Lake Mendota water temperature profiles and the first "social science" data set on boat counts from the Yahara locks on Lake Mendota.

Metadata from the online data sets on the World Wide Web (WWW) were incorporated into an Oracle table. A WWW user interface was created to allow authorized researchers to enter additional metadata directly into the Oracle database. That project led to the development of a data entry program to allow web form data entry into any Oracle table controlled by information placed in a "data entry control" table. This program utilizes the Oracle Web Server version 2's pl/sql "cartridge". We want to centralize business rules in the database (as triggers and constraints) and move away from distributed entry systems where small subsets of data are entered via a program such as Excel into many separate files. Such systems have versioning problems, use easily corrupted, quickly obsolete proprietary binary file formats, and cannot enforce data integrity as well as in Oracle.

Spatial data were made available through the NTL home page on the WWW together with supporting information. Work is underway to consolidate all NTL spatial data coverages onto the Sun workstation.

Hardware and software:

The network administrator has recently replaced the Novell file server with an NT server. On the Solaris platform, Stubbs has written perl programs to do custom by depth and by date interpolation and has used MatLab to produce color contour plots of water temperature (or chlorophyll) by depth and date. Researchers are finding these plots useful. Stubbs has installed the freeware program Samba on Solaris and has found it to be very convenient for win95/NT users to access files on the Sun using the smb protocol. No new software needs to be installed on the client machines to do this.

Intersite information management activities:

The two NTL data management staff attended the CLIMSTAN meeting (LTER Climate Committee and data management representatives) in Albuquerque, NM in September, 1997. Stubbs has continued CLIMDB prototype development at our site during the past year. The next step that is pending will be the migration of the CLIMDB prototype to the LTER Network Office. This migration will involve transferring the database tables from an Oracle database on a Unix platform to MS SQL on an NT Server.

Palmer Station (PAL): Karen Baker

The Palmer LTER information management efforts continue to include data preparation for database entry, data synthesis for publication, field support and intersite activities. Considerable time was dedicated to data review and preparation for entering in the database maintaining compliance with the NSF data policy.

A variety of meetings, workshops, presentations and publication activities have been supported with web based efforts. The Palmer LTER web pages were expanded to include Palmer LTER as well as Network LTER data management with projects, meetings, surveys and milestones subsections as well as presentations.

Data Management synthesis efforts included preparation of draft Antarctic Journal articles for the recently completed season (Palmer LTER: Annual January Cruises for 1998; Annual Season Sampling at Palmer Station November 1997-March 1998) and final preparation of last year's reports accepted for publication. An analysis of satellite sea ice data presenting a standard description of the timing and magnitude of ice events was completed (RSmith, KBaker, SStammerjohn, Exploring Sea Ice Indexes for Polar Ecosystem Studies, Bioscience 48:83-93). Group co-ordination through web based document sharing was successful with the multi-author publication (Marine Ecosystems Sensitivity to Historical Climate Change: Antarctic Peninsula, Bioscience accepted) resulting from a Palmer LTER paleohistory workshop in August 1997.

Summary data tables were created for meetings such as the October 1997 Palmer LTER steering committee meeting. An updated data management overview was presented at a May 1998 NSF Palmer LTER program review. Past presentations at the August 1997 Data and Information Management in the Ecological Sciences Workshop in Albuquerque were made available online and two articles (Technological Underpinnings: Software; Palmer LTER Information Management) have been prepared for publication as part of the Workshop electronic book.

Support for a summer half time student focuses exploration of system integration of a recently purchased NT system with scanner, data presentation methods, and relational database prototypes.

Field support involved weekly real-time ftp of field data for screening and archive made possible with the several hour per day satellite link to Palmer station. As a result, study maps and eventlogs were available upon study completion. Also, weather data collection was reviewed in preparation for a June 1998 Palmer Area Users Meeting covering consideration of local weather operation instrument upgrades, weather observations provided for the Global Telecommunications System, and maintenance of three Automatic Weather Stations.

Intersite activities included participation in a variety of data management meetings and LTER science meetings. The LTER Coordinating Committee (CC) Meeting October 1997 was hosted at UCSB. The meeting information was maintained online. Notes on host tasks and registration forms were added afterward for use of future hosts. Logistics support was provided for DataTask conference calls and a DataTask meeting February 1998 hosted by Andrews LTER in conjunction with Northwest Alliance for Computational Science and Engineering (NACSE). Participation in the LTER Climate Committee meeting October 1997 resulted in a presentation at the October 1997 LTER CC "The Climate Prototype and the LTER Network Information System". Consideration given to presenting uniform site information views has resulted in development of an online SiteDB prototype with implementation by Darrell Blodgett (BNZ). The prototype was reviewed by DataTask and will be considered at the 1998 data manager meeting.

Plum Island Sound (PIE): Hap Garritt

The Plum Island Ecosystems (PIE) LTER is in the process of shifting our focus from previously funded LMER research to comply with the guidelines of becoming an LTER site. The shift from the LMER to the LTER has so far been relatively painless as much of our LMER infrastructure is easily adaptable to LTER.

The primary focus of PIE research will be to address the following question and hypotheses: How will trophic structure and primary and secondary productivity in estuaries be affected by changes in organic matter, nutrient and water fluxes caused by changing land cover, climate and sea level? Hypothesis 1. The interaction of inorganic nutrients with the quantity and quality of organic carbon and organic nitrogen plays an important role in determining the trophic structure, production and efficiency of estuarine food webs. Hypothesis 2. The variability in land, ocean and atmospheric forcing is a key component determining the fate of allochthonous and autochthonous materials and the location and magnitude of primary and secondary productivity.

PIE officially started on July 1, 1998 but to avoid wasting the summer field season we actually started field work in May 1998, focusing on characterizing various marsh sites and tidal creeks for future long term experiments. Through the help of Randy Carlson, VCR, we have established the beginnings of a GPS benchmark network which will be useful for determining elevations of marsh and tidal water throughout the Plum Island Sound estuary. We plan on installing a meteorological station and tidal water monitoring stations in the Fall of 1998. Establishment of marsh surface sediment erosion table networks are also planned for the near future. Data and information management of PIE is a continuous process. We will be converting the old PISCES LMER home page to a new PIE LTER home page and will be incorporating LTER specific guidelines for database access.

Sevilleta (SEV): Gregg MacKeigan

<none>

Short Grass Steppe (SGS): Chris Wasser

In the past year, we have continued to make more datasets available via our website, although there is still a backlog of datasets waiting for QA/QC. This process of cleaning and posting datasets continues to be the highest priority for SGS information management. It is our goal to be in complete compliance with NSF's stated protocol of having datasets available via the web in two years or less.

Another area of activity is the complete redesign and development of our website. We have converted our website into Frontpage98 and are currently conducting extensive error checking on our HTML, links, and text. We plan to have our new website online by the end of 1998. We will also be converting our webserver to an NT machine running MS internet information server. We are hoping that this tighter integration between site management and server software will facilitate easier management of our website.

Plans for this year include moving the SGS LTER database out of Oracle and into MS Access. We have found ORACLE to be too much of an administrative headache and to be honest we don't really need all the power of a database like Oracle. We have selected Access, since it meets our data and information needs and is very simple to manage and administer. With its integration with the above server products, we hope to add more database applications and functionality to our website in the coming year. We plan to have the data migration and web programming completed by early March 1999.

With respect to GIS data, we have completed a data rescue project of quadrat data from Hays Kansas. This was a mammoth project involving countless hours of student digitizing, cataloguing, and QA/QC. Currently all maps are completed and we are in the process of developing and executing an archival process for these data. In addition, we now have complete soils GIS data for the entire Pawnee National Grasslands.

Virginia Coastal Reserve (VCR): John H. Porter

Information management at the VCR/LTER experienced continued growth and consolidation during the past year. The VCR LTER Information Management System was extensively revamped during late 1996 and early 1997. A major step was taken to allow PI's to directly input and modify metadata (documentation) using online forms on the WWW. The metadata is maintained in a relational database (MSQL) running on a UNIX platform.

The VCR LTER WWW server transferred 807,458 files during May 1997-May 1998. This included an average of 939 requests per day for WWW pages (not graphics). There were a total of 66,520 different computers accessing the system during that period. Educational users made up the largest single block of users, comprising 46% of overall requests and 47% of all bytes downloaded. 110 different countries were identified among the user-base.

On the LTER-Network level, the VCR/LTER was active in the conversion of the prototype data catalog to a fully-operational data catalog at the LTER Network Office (http://www.lternet.edu/DTOC).

It has been a busy year for presentations on LTER Information Management with talks presented to The Ecological Society of America (with Don Henshaw and James Brunt), the Biodiversity and Ecosystem Informatics Working Group (a Federal task group), the International Congress of Ecology, and the IEEE Metadata Meeting. In addition we attended the CODATA meeting, the 4th Latin American LTER workshop and an LTER display to congressional aides and representatives.

In the coming year, with a supplement from NSF we will be strengthening our program for archiving physical

samples. This activity will be conducted in partnership with the Virginia Museum of Natural History. We will be cataloging existing and new samples and placing the catalog online via the WWW. Samples will include a variety of tissue samples (suitable for genetic and stable isotope analyses), a herbarium and soil and water samples. We also plan to do additional investigations on uses of XML to help provide additional structure to our text documents.

C. 1998 Web Survey (K. Baker (PAL))

The table below summarizes the diversity of approaches to presenting information at the top level or home page of each LTER site. Web page design involves a variety of decisions about how to organize information in an optimal manner. There are no right answers since research, audience and breadth of information varies considerably, but viewing the pages in an overview summary provides a context within which to consider effective strategies for conveying information. This survey deals only with the initial or top level page. Thus a topic may be listed as missing in the following table, but the information may exist on pages below the single top page.

Information that is generally common across all LTER sites divides into ten main categories: background, publication, presentation, research, data management, data, contacts, links, events, and outreach. Some sites have as few as 6-10 top level home page categories while others include up to 22-25. A survey such as this provides a mechanism for the gathering and sharing of information across distributed sites.

http://www.icess.ucsb.edu/lter/dm/surveys/webdesign/webdesign98.html

D. 1998 LTER Software Survey (K. Baker (PAL))

The annual LTER software survey records the diversity of software used within the LTER community. Results from 1991 through 1997 have been discussed in previous publications:

*J.Porter, R.Nottrott, K.Baker, 1996. Tools for Managing Ecological Data, Proceedings of the Eco-Informa Workshop, Global Networks for Environmental Information, Lake Buena Vista, FL, 11:87-92.

*K.S.Baker, 1998. Technological Underpinnings: Software, Data and Information Management in the Ecological Sciences: A Resource Guide. 1998. W.Michener, J.Porter, S.Stafford (Editors)

The following table has been updated for 1998 by each site's data manager.

http://www.icess.ucsb.edu/lter/dm/surveys/sitesoftware/98lter