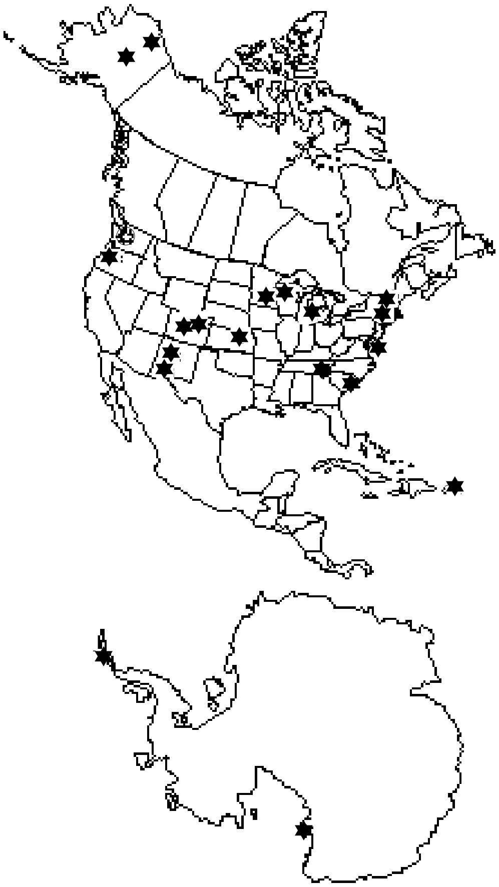
Proceedings of the 1994 LTER Data Management Workshop



Editors: Rick Ingersoll and James Brunt

Workshop Organizing Committee: Barbara Benson, Karen Baker, Rick Ingersoll, Rudolf Nottrott, Caroline Bledsoe

Participants and Contributors: K. Baker, B. Benson, C. Bledsoe, D. Blodgett, J. Briggs, J. Brunt, G. Calabria, T. Callahan, S. Chapal, H. Chinn, J. Faundeen, M. Folk, J. Frew, D. Fulker, K. Gardels, M. Gentry, J. Gosz, D. Gould, B. Gritton, M. Harmon, J.Hastings, J. Helly, D. Henshaw, M. Holland, R. Ingersoll, J. Jefferson, T. Kirchner, M. Klopsch, L. Krievs, J. Laundre, K. La Fleur, C. Lehman, R. Lent, G. Lienkaemper, R. MacArthur, A. McKee, D. Mark, L. May, E. Melendez, W. Michener, M. Murillo, B. Nolen, R. Nottrott, J. Porter, J. Quinn, B. Shepanek, G. Shore, S. Stafford, N. Tosta, R. Turner, C. Veen, L. Walstad

Long-Term Ecological Research Network (LTER)

Long-Term Ecological Research Network
College of Forest Resources, AR-10
University of Washington
Seattle, Washington 98195
World Wide Web http://LTERnet.edu/

For instructions on access to electronic information on LTERnet (including this report) send e-mail to
Overview@LTERent.edu   or
Helper@LTERnet.edu

Proceedings of the 1994 LTER Data Management Workshop

Table of Contents

Proceedings of the 1994 LTER Data Management Workshop

1 February 1995

Editors: Rick Ingersoll and James Brunt
Workshop Organizing Committee: Barbara Benson, Karen Baker, Rick Ingersoll, Rudolf Nottrott, Caroline Bledsoe
Participants and Contributors: K. Baker, B. Benson, C. Bledsoe, D. Blodgett, J. Briggs, J. Brunt, G. Calabria, T. Callahan, S. Chapal, H. Chinn, J. Faundeen, M. Folk, J. Frew, D. Fulker, K. Gardels, M. Gentry, J. Gosz, D. Gould, B. Gritton, M. Harmon, J.Hastings, J. Helly, D. Henshaw, M. Holland, R. Ingersoll, J. Jefferson, T. Kirchner, M. Klopsch, L. Krievs, J. Laundre, K. La Fleur, C. Lehman, R. Lent, G. Lienkaemper, R. MacArthur, A. McKee, D. Mark, L. May, E. Melendez, W. Michener, M. Murillo, B. Nolen, R. Nottrott, J. Porter, J. Quinn, B. Shepanek, G. Shore, S. Stafford, N. Tosta, R. Turner, C. Veen, L. Walstad

1.0 Executive Summary

The 1994 Long-Term Ecological Research Data Managers Workshop was held September 21-24 in Seattle, Washington. The format was expanded from that of previous meetings in order to foster interaction with non-LTER scientists and information managers facing challenges similar to those of the LTER data managers. A total of 25 data managers and other interested scientists attended the LTER business portion of the workshop (21-22 September), including at least one representative from each LTER site. This portion of the workshop focused on the continued development of strategies and standards for the facilitation of inter-site research, and included working groups on metadata standards, information system development, data publication, and technological needs.

An additional 23 individuals, representing a variety of educational, governmental, and other organizations, participated in the "open" session of the workshop (23-24 September). The open workshop focused on two major themes, management of spatial data and inter-site data access, and featured invited speakers as well as working groups addressing various aspects of those themes.

1.1 Working Groups: LTER Business Session

1. **Metadata standards.** Participants moved closer to completion of a set of metadata standards that represent a step in improved *capability for inter-site analyses and research*. The effort builds on the previous year's adoption of a set of minimum metadata components and incorporates appropriate elements of developing federal and non-governmental organization standards.[Section 2.1; pp 5-8]
2. **Revised MSI.** Participants have drafted and continue to refine a document defining *"recommended technological capabilities"* (RTC) for LTER sites that will replace the now dated MSI (minimum standard installation). This document will provide general guidelines for necessary technologies and specific examples of implementations and costs, including personnel.[Section 2.2; pp 8-11]
3. **LTER information system.** A working *group* has been established *to pursue the design and development* of an LTER-wide information system. The design of such a system would incorporate components for the facilitating inter-site research as well providing an information server to non-LTER entities.[Section 2.3; pp 11-15]
4. **Data publication.** A working *group* has been established *to develop mechanisms for* data publication. The objective is to provide a positive incentive for investigators to develop, maintain, and make accessible adequately documented datasets. Suggested mechanisms include development and distribution of data citation standards, as well as creation of a *prototype peer-reviewed data journal*.[Section 2.4; pp 15-16]

1.2 Activity Reports: LTER Business Session

1. **All-Site Bibliography.** The bibliography *continues to expand* --- both in terms of individual entry content (e.g., addition of abstracts) and number of entries. *Increased usage* of the bibliography over the Internet was *documented*.[Section 3.1; pp 16-17]
2. **Data management committees structure.** The *organizational structure* of the LTER data managers as a whole was *described*, including committee functions, membership, and meeting frequency.[Section 3.2; pp 17-18]
3. **On-line access to LTER data.** Remote accessibility to LTER data has been summarized by topic and site. Tremendous *progress* has been *made in recent months* and indications are that such progress *will continue*.[Section 3.3; pp 18-19]
4. **Connectivity Committee.** The LTER connectivity team *conducted electronic communications workshops* for NSF's Division of Environmental Biology in Arlington, VA, INTECOL meetings in Manchester, UK, and ILTER meetings in Rothamsted, UK.[Section 3.4; p 20]
5. **Core Data Set Catalog.** The current status of the *on-line version* of the catalog was presented with details provided regarding the manner in which the catalog can be updated and *features* that *facilitate access to* the desired *information*. The new prototype *image catalog* was *described* briefly as well.[Section 3.5; pp20-21]

1.3 Working Groups: Management of Spatial Data Session

1. **Network of networks.** The group *discussed means by which communication* among information managers *could be enhanced*, a sense of community for communication could be provided, and recommendations related to these.[Section 4.1; p 21]
2. **Spatial metadata standards development.** The group *identified* the *requirements* of standards, key metadata *elements, and means* by which development of standards could be effected.[Section 4.2; pp 22-23]
3. **Spatial data exchange.** The group *identified* probable exchange *"players", topics* requiring additional discussion (e.g., legal ramifications), possible *formats* for exchange (as well as issues associated with format), *and recommended* development of metadata *standards*.[Section 4.3; pp 23-24]
4. **User access and image catalogs.** The group determined that a catalog *should include metadata*, should *be searchable* (possibly by different mechanisms), would be *useful*, and was *feasible* for the LTER sites using extant Internet tools.[Section 4.4; p 24]
5. **Proprietary issues.** The group acknowledged that an increased demand for on-line accessibility of ecological data would necessitate ongoing discussion of such issues. The focus was on *means by which fears* of the community *could be allayed and rewards* for improved data accessibility could be *implemented*.[Section 4.5; pp 24-25]

1.4 Working Groups: Inter-site Data Access Session

1. **Fundable data management research topics.** The group *identified* both *technical* (e.g., "glue" to facilitate use of multiple tools) *and institutional* (e.g., interagency methods for coordinated entry/use of databases) *topics*, as well as related funding issues (e.g., investigator-initiated vs. programmatic).[Section 5.1; pp 25-26]
2. **Metadata standards and exchange.** The group *focused on* development of a *standard* for metadata *content* and acknowledged the necessity of coordination with other organizations (e.g., FLED, FGDC). Metadata standards and standards necessary for exchange were discussed and *action items were identified* for development of those standards.[Section 5.2; pp 26-27]
3. **On-line information system implementation.** The group acknowledged that the collaborative nature of *LTER research requires easy accessibility to information*. The *role that software* such as Gopher and Mosaic *can play* in facilitation of accessibility were the primary focus, and information was provided (i.e., URLs) that will enable use, implementation, and maintenance of the client and server versions of these products.[Section 5.3; pp 27-28]
4. **Network tool/software interfaces.** URLs for a number of *potentially useful* (for data managers, working groups, scientists, etc.) *tools* were *listed* and those sources will be made available on LTERnet.[Section 5.4; pp 28-29]

1.5 Invited Presentation Summaries

1. **Data management and storage: today and tomorrow.** *Jim Frew*'s presentation *focused on* (1) *technological trends* affecting data management *and* (2) *the DBMS-centric data management paradigm*. The former topic addressed processor technology trends, storage technology trends, and network technology trends; some recommendations were provided based upon these trends. The latter topic was summarized and examples of solutions were provided.[Section 6.1; pp 29-37]
2. **Management of geographical data.** *Kenn Gardels noted that hybrid systems/tools will be needed* for effective management and analysis of "geodata". In addition, some degree of *standardization* within the realm of data modeling will be *necessary* to facilitate data exchange. A thorough summary of geodata modeling issues was provided, as were a number of relevant URLs.[Section 6.2; pp 37-43]
3. **The National Spatial Data Infrastructure.** *Nancy Tosta reported* that the Federal Geographic Data Committee is promoting the *development of NSDI through* (1) establishment of a clearinghouse to facilitate *access to,* (2) conceptualization and development of a digital framework dataset that will minimize redundancy and facilitate *integration and use of, and* (3) *development of standards for geospatial data*. Development of NSDI *will require cooperation* among different levels of government, as well as public and private sector entities.[Section 6.3; pp 44-45]
4. **Geospatial data acquisition, quality, and metadata.** *Bill Michener stated* that many of the constraints associated with use of GIS had been removed and, as a consequence, *increasingly important issues* in the future will be *how to best identify, acquire, manage, and analyze geospatial data* that are timely, relevant, and of sufficient quality. Some *potential technical solutions were provided,* although it was noted that technology alone cannot solve all of the problems.[Section 6.4; pp 45-52]
5. **Models, formats, and tools for inter-site data exchange.** *Dave Fulker sketched* the Unidata Program Center and *two categories of software offered --- analysis/display and data management*. From the second category, the *netCDF* was used *to illustrate how a data model* and portable software tools to perform the storage/retrieval fucntions of the model *can be more effective than standard formats,* per se, to facilitate inter-site data exchange. [Section 6.5; pp 52-58]
6. **Integration of user interfaces and data formats.** *Mike Folk provided examples of modern user interfaces* (e.g., clients, GIS applications) and standard data formats (e.g., HDF, netCDF). An *evolutionary view of access* from the "snail mail" to the present client-server models was *outlined* and several client/server examples were provided.[Section 6.6; pp 58-60]
7. **Challenges/opportunities for information managers: The NSF/BIR perspective.** *Susan Stafford described* the nature and *impact of* the current *information explosion/technological revolution* within/upon the biological sciences. The concept of the "*collaboratory*" was introduced within that context. NSF's *Division of Biological Instrumentation and Resources will play a major role* in the development of the infrastructure for such a laboratory, particularly through the (1) Database Activities Program and (2) Research Training Groups.[Section 6.7; pp 60-61]

2.0 Working Group Reports: LTER Business Session

2.1 Documentation Standards for Data Exchange

Committee: Tom Kirchner (CPR), Harvey Chinn (NET), Don Henshaw (AND), John Porter (VCR)

Proposed content. The discussion of metadata standards continued at the 1994 meeting of the LTER data managers. The interest in formalizing a standard arose from discussions (at the previous two meetings) on methods for facilitating the exchange of data between sites and to other interested researchers. We are proposing that the metadata for any site can include any information needed to facilitate data management at that site, but that a minimum set of information be included in order to facilitate data exchange. This list of metadata items is independent of implementation, although we expect to implement a standard format to be used when exporting data. The export standard for metadata will be based on the format developed by Tom Kirchner for use at the CPR LTER site.

The proposed content for metadata lists items considered necessary for understanding and using a set of data. These items have been classified as information about the dataset as a whole, and information about the variables (attributes) contained in the dataset. The metadata for variables is intended to be applied only to data stored as fixed field records within ASCII files. Similar kinds of information will be needed to describe binary data, or data in variable-length delimited ASCII files, such as DIF files. Spatial metadata are excluded because we are recommending the use of the standard described in “Content Standards for Digital Geospatial Metadata” (Federal Geographic Data Committee 1994). Items marked **1** were identified by the task group as being recommended but optional, whereas items marked **2** were identified as being optional.

Names entirely capitalized are recommended by the Federal Geographic Data Committee and National Biological Survey. The other names are in use or are proposed for use in metadata documentation from the LTER sites.

Dataset information

- DATASET IDENTITY - name or title by which the dataset is known (type text, domain free text)

  Type of data [fixed field, delimited field, binary, image, etc.)

- ORIGINATOR - name of the organization(s) or individual(s) that developed the dataset. If the names of editors or compilers are provided, they must be followed by (ed.) or (comp.)

  Originator address

  Originator voice telephone

  Originator facsimile telephone

  Originator electronic mail address

- TIME PERIOD OF CONTENT (start and stop dates)

  Dissemination restrictions

- AVAILABLE TIME PERIOD or RELEASE DATE (NBS)

- DATASET DESCRIPTION - a description of the (spatial) dataset including its intended use and limitations (type text, domain free text)
- CONTACT INFORMATION
  - CONTACT PERSON
    - CONTACT MAIL ADDRESS
    - CONTACT VOICE TELEPHONE
    - CONTACT FACSIMILE TELEPHONE
    - CONTACT ELECTRONIC MAIL ADDRESS
- KEYWORDS -words or phrases summarizing some aspect of the dataset (type compound)
  - THEME KEYWORD (required)
  - PLACE
  - STRATUM
  - TEMPORAL
- IDENTIFICATION CODE[1] -unique item or stock code by which the item could be ordered or the full path name to the file (type text, domain free text, N/A, UKNOWN)
- DATASET CITATION or CITATION_INFORMATION[1] (preferred)
- SPATIAL DOMAIN[1] (bounding coordinates, or arbitrary polygon)
  - Sample storage[1] (Information on storage of samples)
- RELATED DOCUMENTS[1]- Reference citations and location information

Attribute (variable) documentation

- ATTRIBUTE LABEL
  - Type [integer, floating point, character/string]
  - Format
    - Start column
    - End column
    - Optional number of decimal places
- ATTRIBUTE DEFINITION
- ATTRIBUTE UNITS OF MEASUREMENT
- ATTRIBUTE DOMAIN VALUES - CODE SET DOMAIN
  - ENUMERATED DOMAIN VALUE DEFINITION (list of coded values)
  - RANGE DOMAIN
    - RANGE_DOMAIN_MINIMUM
    - RANGE_DOMAIN_MAXIMUM
  - Missing value codes[1] (should be required if such codes can appear in the data)
  - Precision[2]

Methods of collection

Other metadata attributes. The following items were derived from a review of metadata requirements from the LTER sites. Two additional categories of metadata were identified: information that primarily describes the project under which data were collected, and metadata that describes characteristics of the sites from which the data were obtained.

- Project description

  Project Title

  Date commenced

  Date terminated or Expected duration

  Objectives

  Abstract

  Source of funding

  Principal investigator

  Additional investigators

  Responsible investigators/technicians/supervisor

- Site description

  Site type (terrestrial, aquatic stream, aquatic lake, etc.)

  Watersheds

  Permanent plots

  Habitat

  Soil

  Slope/aspect

  Terrain/physiography

  Geology/lithology

  Hydrology

  Size

  History

  Elevation

  Climate

- Dataset description

  Dataset title

  Number of records

  Data form used

  Location of completed data forms

  Literature

Method of recording

Associated computer accounts

Comments

LTER core area

Dataset files (subsets)

Citation

Treatment of data (programs used in analysis, plus reference/link to their metadata)

Date of last review

Date of last entry

Researcher review status

Taxa/functional group

Statistical analyses

Supporting datasets

Other supporting materials

Maps

Aerial photos

Digital images

GPS

Use history

Request histories

Update histories

The recommended list of metadata should be reviewed by the data managers and other interested people at the sites. The following things need to be done:

1. Consensus must be achieved on the set of items.
2. The names for the items must be standardized.
3. Definitions must be provided for all items.
4. Where items must be assigned values from a specific domain, such as Type of data (=fixed format, binary, delimited ASCII, etc.), the values that can be assigned must be identified and defined.

2.2 Recommended Technological Capabilities for Computational Environments in Long-Term Ecological Research (LTER) Projects

Committee: John Briggs (KNZ), James Brunt (SEV), Mark Klopsch (AND)

Rationale: The rationale for this document was based upon the need to update the LTER worksheet of November 1988 --- MSI (Minimum Standard Installation) Technological Capabilities at All Sites of the LTER Network. That document was extremely useful in establishing minimum levels of technological capability in order for the LTER network to accomplish several inter-site goals and for each of the LTER sites to reach a minimum level with respect to hardware and software. More importantly, it identified three major computational technologies that the LTER sites needed to accomplish many of these goals. They were: a) GIS facilities (including hardware, software, and technical staff), b) access to both local- and wide-area networks, and c) high-capacity, archival mass storage systems.

The MSI was successful in that it led to a strengthening of the technological capabilities of many of the LTER sites, as well as facilitating development of the LTER network. The status of the LTER network goals was reviewed in an earlier document (LTER Network Office Publication No. 12). In addition, as stated in the "Ten-Year Review of the National Science Foundation Long-Term Ecological Research (LTER) Program", the "MSI elements include the common hardware and software elements necessary for the computer networking system that links eighteen sites as well as standardization of other computer applications such as GIS". In order to enhance the research capabilities of each site and ensure continued growth as a network, we propose recommended technological capabilities (RTC) with respect to 3 major areas. We believe that this document can:

- 1) Facilitate the LTER network response to the challenges raised in the "Ten-Year Review of the National Science Foundation Long-Term Ecological Research (LTER) Program" report prepared by a review committee, co-chaired by Paul G. Risser and Jane Lubchenco. New technology will play a vital role in the proposed expanded and enhanced LTER network. It is critical to have the necessary technology if the LTER sites are going to integrate and synthesize results both within and among sites and seek ways to generalize these results over broader spatial and temporal scales as suggested by the ten-year review.
- 2) Provide information to all LTER sites that will allow them to make decisions concerning technological capabilities necessary to become a functional component of the LTER network and to further enhance the continued development of the LTER network. In addition, this document can also provide such information to the larger ecological community.
- 3) Provide NSF with the necessary information to adequately evaluate requests from and allocate funds to LTER sites.

Definition of RTC: The following text describes recommended technological capabilities (RTC) for LTER sites in several areas of information technology. Adequate development and integration in all three of these closely inter-related areas is required in order to create a functional computational environment for the management and analysis of scientific information.

A. SCIENTIFIC DATABASE MANAGEMENT SYSTEM. A combination of hardware, software, and personnel to have a functional scientific information system.

B. DIRECT HIGH SPEED ACCESS TO THE INTERNET. This should include access from the institution, the desktop, and the field.

C. SPATIAL, VISUAL, STATISTICAL, AND MODELING LAB. This should include GIS/RS capabilities, a wide range of statistical and visualization software and a combination of hardware, software, and technical personnel necessary to promote ecological modeling.

A. SCIENTIFIC DATABASE MANAGEMENT SYSTEM.

The purpose of a scientific database management system at an LTER site (or any other ecological research site) is to insure the high quality, adequate security, and accessibility of scientific data, as well as to promote the synthesis and integration of long-term datasets originating from current and past research. In addition, the system must be sufficiently flexible to (1) facilitate integration of research with other LTER sites, (2) handle new and innovative research programs that are being encouraged as the LTER sites expand their scope, and (3) encourage and facilitate the use of the data by both site and outside scientists.

To accomplish these tasks, we have identified three major technological areas that a site should address to have an adequate scientific database management system.

1) Computers, Software, and Peripherals: In addition to supporting local database activities, these components must be sufficient to provide easy access (*in an integrated fashion across LTER sites*) to adequately documented long-term datasets, both locally and remotely through the Internet. While browsing capabilities are useful for metadata, the database system should also support remote logical queries on the full data using industry-standard SQL. Many sites may require increased processing power and an upgraded database management system (DBMS)

2) Storage Management: Although a storage management system is actually implied in item 1, the need to provide full on-line or near on-line access to datasets; adequate disk space for data management, GIS, and remote sensing analyses; modeling; consistent back-ups; and archival storage warrant special attention. A strategy might involve the use of multiple disk arrays (RAIDS) for on-line access and near on-line access via the use of tape arrays. It is expected that a site would use a variety of electronic media for their back-ups and for archival purposes. In addition, a site should maintain off-site copies of “back-ups” for additional security.

3) Personnel: In order for an LTER site to meet its goals, as well as those of the network, it must be adequately staffed.

B. DIRECT HIGH SPEED ACCESS TO THE INTERNET.

The completion of a high-speed network infrastructure is essential for cross-site data synthesis activities in a distributed system. In addition, the further development of a network-wide information and analysis system will hinge on the connectivity of the computers involved in the management and analysis of LTER data. Technological improvements in this area will soon allow teleconferencing and thus tele-collaboration. Sites should seek full implementation of the recommendations put forth in “LTER Network Office Publication No.7; Internet Connectivity in the Long-Term Ecological Research Network” to increase the availability of high-speed connections at all levels of computers involved in LTER, institutions, project computer systems, individual workstations, and field sites.

C. SPATIAL, VISUAL, STATISTICAL, AND MODELING LAB.

It is critical that all of the LTER sites have the tools and personnel necessary to integrate and synthesize research results (both within and among sites) and seek means to generalize those observations over broader spatial scales (as suggested by the 10-year review of LTER). The emphasis at each site will vary depending on stage of development and research direction. The MSI worksheet, for the most part, emphasized GIS tools; GIS development (including the use of global positioning systems) should continue, but the addition of visualization tools (remote sensing capabilities and data analysis), statistical software packages and modeling tools (both hardware and software) should also be emphasized. Multiple input devices for data capture (scanners, digitizers, digital cameras, etc.) and output devices (plotters, color printers, film recorders, etc.) are needed. Support for this capability in the form of trained personnel is equally critical and will facilitate synthetic efforts.

Summary. We envision that a 1994 "state-of-the-art" laboratory would include GIS software that fully integrates vector/raster databases, statistical software packages that include spatial analysis as well as more "traditional" approaches, and data visualization software that allows scientists to more easily evaluate complex datasets. Full integration of these tools would aid in the use of models for synthetic activities.

The attached appendix includes examples of computational systems in use as well as a generic system that could be used as a template for a start-up system.

2.3 Development of an LTER Information System

Committee: Jordan Hastings (MCM), Karen Baker (PAL), Barbara Benson (NTL), Darrell Blodgett (BNZ), Gil Calabria (CWT), Harvey Chinn (NET), Tom Kirchner (CPR), Lolita Krievs (KBS), Barbara Nolen (JRN), John Porter (VCR), Greg Shore (SEV)

A working group met to discuss the development of an "information system" at the network level, since individual sites already have systems. Participants all recognized that such a network system would be a large, complex undertaking. From initial discussion, however, it was clear that people understood the task in very different ways.

Accordingly, in the short time available, the group agreed to try brainstorming on 2 topics:

1. Technical Issues, which would need to be addressed in any system plan, e.g.,

    Capabilities

    Interface with users

    Performance

    Privacy/security

    Staffing

2. Procedural steps which might be taken to bring such a system closer to reality, e.g.,

    Requirements analysis

    Planning workshop

    External study/recommendations

    Prototype implementations

From these discussions emerged consensus on a quasi-formal, central (v. the current ad hoc, distributed) approach to information systems planning, encompassing successive refinement of a vision statement into a strategic plan, into an implementation plan, and finally an implementation. The co-involvement and support of the LTER PI community was deemed important throughout this process. Specifically, the group recommended the following actions:

1. Drafting of a concise vision statement (see below), to be reviewed through DataTask and DMan, then presented to the LTER Coordinating Committee to test support for further development.
2. Request for internal funds (under the new LTERnet cooperative agreement) to hold a systems planning workshop in winter/spring 1995, the outcome of which would be a draft strategic plan to provide the framework for implementation planning.
3. Development of a formal implementation plan by autumn 1995 (if possible) for review again at next year's Coordinating Committee meeting, which would identify sub-system components to be developed as prototypes.
4. Proposal(s) to NSF (particularly the Database Activities Program in the Biological Instrumentation Resources Division) in November 1995 --- the next possible funding cycle --- with the hope of beginning actual implementation work in summer 1996.

VISION: INTEGRATED DATA & INFORMATION SYSTEM FOR THE LTER NETWORK

Background. In their first decade, the Long-Term Ecological Research (LTER) sites have generally distinguished themselves in traditional, site-specific, ecological research. Data and information systems (DIS) adapted to the sites' individual scientific and sociological needs were instrumental in this success. With the mounting demands for multi-site, synthetic work, however, a new level of DIS capabilities is required, one which integrates seamlessly across the LTER network.

Toward this end, the LTER Network Office (LTERnet) has begun implementation, over the past several years, of a DIS that holds great promise. This system currently provides information sharing in numerous venues: electronic mail (email), bulletin board systems (BBS), and browse/ lookup tools (Gopher, Mosaic) across the widely-distributed community of LTER scientists and support personnel. File transfer protocol (FTP) is provided as well. The bulk of material currently being shared on LTERnet is textual.

The LTERnet system also has the capability to maintain and share binary data in many forms. These forms include, but are not limited to databases, spreadsheets, graphics/images, audio, video, etc. To date, however, such non-textual data sharing has been hampered by lack of standard methods for identification and description of the data, i.e., the metadata. Ecological data are inherently complex, and thus "rich" in such metadata. Although some free-format, textual metadata sets exist, even within LTERnet, these are non-standardized (and probably not standardizable); they are of greatest benefit to knowledgeable human readers.

To facilitate actual data sharing in the larger LTER research community, and especially between computers (v. people), a standardized cataloging system for metadata is required. Inevitably, such a metadata catalog includes binary data; consequently it is recursively entangled in the problem it seeks to solve. Indeed, it seems that scientific metadata are best organized in a meta-database! Implementing this level of technology has proved difficult for LTERnet, as it has for many other organizations.

Access to binary data (and metadata) bases is critical to the future of the LTER network nonetheless. As LTERnet matures, and takes on an identifiable, integrated role in other national and international programs, it is clear that the LTER program’s assets are (1) its people, with their wealth of knowledge and experience, and (2) increasingly, its collection of long-term ecological data/information. From the outset, it was understood that LTER data would be accessible to other investigators across space and time. These commitments are now being called.

Opportunity. LTERnet has a signal opportunity to accelerate inter-disciplinary/synthetic and inter-site/comparative research, as well as to solidify its technical investments to date, through development of an integrated DIS. These expanded research objectives require convenient, informed access to large volumes of disparate data across LTER sites. For this, a “distributed” or “federated” DIS is required.

The envisioned LTERnet DIS (LTERDIS) may best be understood by analogy with a modern university library system. Library materials, or holdings, are systematically arranged in physical stacks, and described in electronic catalogs. The stacks are nondescript, fluid facilities for shelving (or otherwise storing) the materials; in LTERDIS, this is analogous to raw disk space. The holdings themselves, which vary widely in size and shape, must be identified by permanently attached external labels, which can be standardized, i.e., filenames. In addition, holdings are accorded entries in the catalog, which serve effectively as extended bibliographic citations, with author, title, subject, keyword(s), physical description, etc. --- the desired meta-database. Most importantly, each catalog entry also cites identifier label(s), which relate back to the actual holding(s) in the stacks via an evolving stack map: in electronic form, a master directory. Retrieval of a holding from the stacks, finally, requires a sequential search of at most a few shelves, i.e., a few disk reads.

Over centuries of librarianship, variations on this basic model have been developed to deal with all the real-world complexities that traditional libraries face: reports, serials, multivolume sets, new editions, duplicate/desk copies, special/reserve collections, branch libraries, etc. The value of having all holdings cataloged conformably is overwhelming, for both librarians and patrons. Again, quite analogous issues --- and perhaps a few new ones --- apply to the envisioned LTERDIS. The key concept, however, is that all LTER sites, and LTERnet, agree on the content and format of a common metadata catalog.

Using LTERDIS, an individual investigator operating from a local-area networked terminal will have immediate access to catalogs and stacks both at the local LTER site and to parallel holdings at other sites. Both traditional bibliographic and geospatial search mechanisms will be provided. To accomplish these services, at least the catalog system must be fully distributed --- with a centralized master --- so that searches beyond the local system are efficient, with minimum demands on remote systems. This arrangement also provides automatic catalog backup for all sites. Holdings of the stacks themselves will be more closely retained at individual sites, although with time common reference materials (such as maps, name lists, taxonomies, etc.) will naturally gravitate to the central LTERDIS facility.

In a traditional library, materials are typically printed in a natural language, which is readily interpretable to patrons (the skill of interpretation having been trained through education). In the electronic library, by contrast, data/information holdings are recorded in abstruse structures which are not generally accessible to patrons, both because humans cannot sense electromechanical recordings directly, and because the structures are dynamic. Instead, intermediary computer software must be provided to “read” the holdings in the stacks. In LTERDIS, the appropriate software “browsers” and “readers” also will be distributed electronically with the holdings, when needed.

LTERDIS provides the forum of connection not only among the individual LTER sites internally, but also to other burgeoning national and international ecological research programs externally, i.e., inter-library facility. In this latter instance, significant computational effort may be required to bridge between different systems, and issues of data integrity and security rise substantially in importance. It is most appropriate to place these responsibilities at the central level, where they can be efficiently administered for mutual benefit.

Approach. Development of LTERDIS, as outlined above, will be a large effort, requiring the cooperation of numerous organizations --- at least 18 LTER sites plus LTERnet --- over several years. To fully succeed, all parties must be pre-agreed to the work, which in turn must be appropriately funded and managed. Equally important, support for operation and maintenance of the system, once developed, including training and support of users, must be provided on a long-term basis. The concurrent, rapid evolution in computer and networking technologies will certainly force changes to the system throughout its design/implementation/maintenance life-cycle, which also must be accommodated. Finally, it is apparent that LTERDIS entails social as well as technical changes for the LTER research community, preparation for which is essential.

To clarify these issues, a formal systems approach is recommended, according to the following schedule:

Nov ’94	Proposal for LTERDIS Requirements Workshop
Mar ’95	System Requirements Workshop *held* Special mid-year workshop of all LTER data managers, together with interested PIs; this will be facilitated by a systems professional and augmented with invited speakers for related agency programs, in the format of the highly successful autumn’94 LTER Data Management Workshop.
Jul ’95	System Requirements Workshop Report *due* Describes technology and sociology, features and benefits of LTERDIS in detail; identifies R&D issues; provides implementation schedule/budget scenario(s), and establishes follow-on O&M baseline.
Sep ’95	Review of Workshop Report at annual LTER Data Management Workshop
Oct ’95	Final review of Workshop Report at Coordinating Committee meeting
Winter ‘95-’96	Proposal(s) prepared pursuant to Workshop Report, as interpreted by data managers’ and Coordinating Committee’s reviews (possibility of small exploratory development projects as well as large infrastructure construction)

Spring’96	Proposal(s) for LTERDIS submitted to NSF
Fall’96	LTERDIS development commences (pursuant to funding)

Conclusion. The LTER program has achieved remarkable results for a collection of independent research sites. LTER’s next ambition, to explore inter-disciplinary/synthetic and inter-site/comparative studies, requires a more systematic approach, however, particularly with regard to data/information management. The environmental research community as a whole is currently grappling with these issues. Thus the LTER program has another leadership opportunity, to formally develop LTERDIS, an integrated data and information management system for the environmental sciences.

2.4 Data Publication

Committee: Clarence Lehman (CDR), Karen Baker (PAL), Mark Harmon (AND), Rick Ingersoll (NWT), Richard Lent (HFR), Barbara Nolen (JRN), Cindy Veen (HBR)

There are few direct incentives for scientists to archive raw data in a widely accessible form. This working group explored one means of providing such incentives --- a journal devoted to the formal publication of data.

The ultimate goal is to make available, across both space and time, selected high-quality data. The data must be usable by scientists who have no access to the original researchers.

The basic strategy is to augment the present reward systems in such a way that scientists are motivated to provide data with adequate information content. Existing rewards are tied to publication in respected journals and, more importantly, to citations of such publications. By extending similar credit for the publication of data, one could take advantage of the reward systems already in place. In addition to being able to cite and track the data, the impact and value of dataset contributions could be assessed. The latter would be similar to what is currently used for traditional publications --- a publication not cited has not been “used” and thus has had little immediate scientific impact.

The specific proposal is to initiate a “data journal” with high standards and adapted to the existing publication and citation rules. The journal would follow strict peer-review procedures and would use the usual volume-issue-page citations of existing journals. Judging from the discussion at the 1994 LTER Data Managers Workshop, this idea seems to be ripe now.

At the workshop, the group enumerated issues and questions. In the following weeks, these were addressed further through electronic mail. A summary of selected questions follows.

- **Who should publish such a journal?** Responses varied, with some individuals suggesting that the LTER group should initiate it, whereas others thought that this task should be left to existing ecological societies. However, even if LTER published the journal, the belief was that it should not be exclusively restricted to LTER data.
- **What should actually be printed (i.e., in hard-copy)?** The consensus was that, with the exception of some “small” datasets, the metadata only should be printed, with both the metadata and the data stored in a read-only format and archived on-line.

- **What type of data should be published?** The consensus was to begin with point tabular data, with subsequent extension to geographical data, simulation-output data, etc., possibly by special groups.
- **What criteria would be used to accept or reject data submitted for publication?** The consensus was that peer-review was a sound approach to assess quality of the data, quality of the metadata, experimental design, adherence to standard formats, etc..
- **Should data be accepted for publication in a data journal *only* after they have been used in an existing refereed publication?** This is probably a good idea for the first datasets accepted, but ultimately should not be a requirement. There are many generally useful datasets which are not in themselves the basis of an exciting refereed publication, but which may form such a basis when combined with other datasets.
- **What is the relative importance, in a data journal, of raw data versus processed and summarized data?** Both are important, but raw data are essential. Conventional journals already provide outlets for summarized data.
- **Is the volume-issue-page citation format sufficient, or should some form of accession numbers be used instead?** A combination may be useful, with a standard citation for the metadata but with special codes for the data. We should coordinate with progressive libraries on this subject.

Other more general questions remain. Would publication in a high-quality data journal come to be respected? That is, would citations to data (indicating that further use was being made of the data) be recognized as a legitimate kind of citation for the authors of the dataset? Would editors of existing journals accept citations to data journals or would they consider them a class of “gray literature”? On the other hand, might editors of existing journals eventually require that supporting data be deposited in accepted data archives. What would be the response of scientists? Would this become a good way to foster additional scientific research and to get additional publications, or would it merely become another task competing for time?

Finally, one proposal for initiating data publication is the following. Conduct an international symposium where generally useful datasets are described --- in terms of composition, past usage, and potential future usage --- and where the datasets themselves are supplied. Each LTER site could submit one dataset for review (but not for guaranteed inclusion), and other groups would be invited as well. Metadata would be published in a proceedings of the symposium with the actual data on a CD (and on-line as well). The proceedings would serve as a prototype for a data journal.

3.0 Activity Reports: LTER Business Session

3.1 All-Site Bibliography (Harvey Chinn - NET)

The All-Site Bibliography (hereafter referred to as the “Bibliography” to make the distinction from a site bibliography) continues to expand as the sites add abstracts and new entries. KNZ is adding abstracts (which are not yet in the Bibliography). PAL has added abstracts for the citations of its LTER-supported research; these have been incorporated into the Bibliography. Seven sites have provided updates over the past ten months. The All-Site Bibliography now contains 9191 citations. Usage of the Bibliography is also increasing. During the time from 1993 May 6 through the end of August 1994, there were 1337 separate searches, with 41% of these occurring in the last four months of the period. Popular search topics include specific people; species and taxonomic groups; community or habitat types; geographical locations; and LTER sites. Of the 746 identifiable searches from U.S. educational institutions, at least 47% came from LTER sites. U.S. government agencies use the bibliography, with the EPA and USDA accounting for 61% of this activity. Reflecting the overall increased usage, the proportion of searches from foreign countries is also increasing. The All-Site Bibliography appears to be functioning as an international resource for ecological work.

Harvey Chinn has programmed a filter that converts the Bibliography's "on-line format" to the export/import format used by Pro-Cite. For those data managers considering switching to a different bibliographic database system at their site, it should be an easy process (for Harvey) to adapt this filter to any import format needed. This may greatly ease the transition to the new system. Please contact Harvey at harvey@lternet.edu for details.

3.2 Data Management Committee Structure (James Brunt - SEV)

Following up on a working group activity from the 1993 meeting, a formal structure of operation for the LTER network data management committee has been adopted. The data management committee (DMC) is a standing committee of the Long-Term Ecological Research (LTER) Network consisting of one representative from each LTER site. It functions to actively develop initiatives and address pertinent issues with regard to the management, exchange, and analysis of ecological data. Decisions are made by consensus. Issues requiring decision are brought before the group as a whole either electronically or in a formal meeting.

The DMC has two standing subcommittees with shared membership: the Data Management Task Force (DMTF) and the Data Management Meeting Organization Committee (DMMOC).

The functions of DMTF are to 1) provide representation for the committee to the Coordinating Committee, 2) ensure continued funding for future meetings, and 3) provide guidance and continuity to the DMC. DMTF will consist of six members, 2 of which are capable of serving as signatory PIs (should the need arise for proposal submission), and the Network Office data manager who will serve ex-officio. One new representative from the DMC membership will be asked to volunteer for a term of 3 years, at that time one of the other members will step down until a fixed-rotation schedule has been established. One of these individuals will be asked to represent the DMC to the Coordinating Committee for a period of 2 years and will be the primary contact person or recognized "chair" of the DMC during that period. That person will have the additional responsibility of notifying the membership of any relevant discussion or directives coming from the LTER Coordinating Committee, Coordinating Committee Chairperson, or Executive Committee. Current members are Barbara Benson, John Briggs, James Brunt, Tom Kirchner, John

Porter (ex-officio), Susan Stafford (ex-officio), and Rudolf Nottrott (ex-officio).

The functions of DMMOC are to 1) organize the meetings of the DMC, and 2) edit and produce the proceedings from those meetings in a timely fashion. DMMOC will consist of six persons, 2 from the DMTF, and 4 additional volunteers from DMC. One of the DMTF members will serve as the “chair” and contact point for the committee. The chairperson may select co-chairs and may ask for additional volunteers from the membership as needed. Current members are Barbara Benson, Karen Baker, Rick Ingersoll, Susan Stafford, and Rudolf Nottrott.

Business meetings are held at least annually and are open to other LTER representation and encouraged especially from the PI and GIS communities. Business meetings are not open to non-LTER representation. In addition, outreach or open meetings may be held focusing on a particular topic in which participation from other groups and agencies is encouraged.

3.3 Status of On-Line Data Accessibility within LTER (Rick Ingersoll - NWT)

The following table summarizes the current availability of data. It should be noted that tremendous progress has been made in this area during the past 6 months and that the evidence suggests continued progress.

Most LTER sites have on-line accessibility to data for site researchers and descriptive metadata for the public. Many datasets corresponding to the latter can be available upon request. This listing, however, is confined to documented (i.e., including metadata) datasets available (without intervention at the site) over the Internet through Gopher, Mosaic, etc.

Note that this table does not reflect quantity or scope of the data, nor does it reflect spatial or temporal extent. Thus, meteorological data could include data collected by more than one source of instrumentation from more than one station for more than one parameter for more than one year. Thus, the number of categories to which a given site has been “attached” in no way corresponds to the amount of data actually available.

An attempt has been made to *generalize* the categories as much as possible. Thus, tree-diameters are considered a proxy for primary production/biomass and are included in that category. Nevertheless, there is often overlap between categories, e.g., fertilization effects and disturbance effects.

No information was available for HFR.

File last updated: 1 February 1995; RCI

Description	Site(s)
aboveground invertebrates	CPR
atmospheric deposition	AND,KNZ
birds	CPR,KNZ
chlorophyll	ARC,PAL
decomposition	AND,CPR,NWT
disturbance effects	ARC,BNZ,CDR,CPR,HBR,NWT
fertilization effects	ARC,BNZ,CDR,NWT
fire history	AND
fish length frequencies	ARC,NTL
GIS coverage maps	AND,ARC,CPR,VCR
global positioning systems	KBS,VCR
litterfall	CPR,HBR,KNZ
meteorology/climatology	AND,ARC,BNZ,CDR,CPR,HBR,LUQ,KBS,KNZ,MCM,NWT,PAL,SEV,VCR
paleoecology	NWT
plant chemistry	ARC,HBR,NWT
plant cover/composition	ARC,CDR,HBR,NWT,KNZ,SEV
plant phenology	ARC,CPR,HBR,KNZ
precipitation chemistry	AND,ARC,HBR
primary production/biomass	AND,ARC,CDR,CPR,HBR,NWT,KNZ
satellite imagery	KNZ,VCR
small mammals	CPR
snow chemistry	NWT
snowpack ablation	NWT
snow physical properties	HBR,NWT
soil carbon	ARC,NWT
soil (water) chemistry	ARC,BNZ,HBR,KNZ,NWT
soil microarthropods/nematodes	CPR,KNZ,NWT
soil moisture/water	AND,ARC,BNZ,CPR,HBR,KNZ,NWT
soil respiration	ARC,BNZ,CPR
soil temperature	AND,ARC,BNZ,CPR,HBR,NWT
stream channel cross-sections	AND
stream flow	AND,ARC,HBR,KNZ,NWT
stream water chemistry	AND,ARC,CWT,HBR,NWT
throughfall	HBR,KNZ
trace gas emissions	ARC,NWT

3.4 Connectivity Committee (James Brunt - SEV)

The Division of Environmental Biology at the National Science Foundation invited the LTER connectivity team to host a workshop/demonstration on “Connectivity”, or “How to Use the Global Electronic Communications Systems for Data Management” purposes. The workshop was intended to be a “realistic” look at connectivity within the LTER network and as a demonstration of how data and information are managed and communicated via networking. The event was aimed at building interagency understanding and use of the Internet as way to facilitate collaboration and foster networks of scientists among the federal agencies. The workshop was held on 16 December 1993 at NSF’s new networked building in Arlington, VA.

The workshop was designed to demonstrate a user-friendly, detailed examination of the means by which one can electronically search, retrieve, manipulate, analyze, and utilize science data from a variety of sources, globally, via the INTERNET. The connectivity team, James Brunt (SEV), Rudolf Nottrott (NET), and John Porter (VCR) demonstrated the use of electronic mail, telnet, FTP, remote computing, Gopher, Khoros, and Mosaic on SunOS and DOS platforms. Attending agency representatives displayed a variety of levels of technical sophistication.

The workshop was then repeated at the INTECOL meetings in Manchester, England and at the ILTER meetings in Rothamsted, England in August. The audience for these workshops were international ecologists interested in information access and connectivity, with many participants being involved in LTER or LTER-like activity. Rudolf demonstrated the prototype of a traveling connectivity kit which included a Unix file server, thin-wire ethernet, and several notebook computers running slip.

3.5 Core Data Set Catalog (Rudolf Nottrott - NET)

The LTER Core Data Set Catalog (hereafter referred to as the “Catalog” to make the distinction from other catalogs) has been integrated into the LTERnet World Wide Web information server (URL **http://LTERnet.edu**) and is accessible from the home page there. The site maps that were part of the original hard-copy version of the Catalog have been scanned and converted to Graphics Interchange Format (GIF) files, and added to the hypertext version of the Catalog. This addition allows convenient viewing of all available Catalog information, including the maps originally produced in hard-copy version. Several sites have recently requested that their Catalog entries be updated. Rather than accomplishing this by incorporating new text files into the LTERnet server and re-indexing the Catalog, we will explore a distributed setup in which a site maintains a local copy and WAIS index of all their entries, and the LTERnet server transparently links to this index on request. If this distributed mechanism works reliably, we will gradually switch over all sites with local Gopher/WAIS capability to maintain their own Catalog entries. Convenient, and transparent central access to the Catalog will still be provided by LTERnet.

Two prototype catalogs have been added to the catalog section on the LTERnet information server, one based on a subset of the Core Data Set Catalog, with direct links to complete metadata and data, and a satellite image catalog with complete metadata and a preview image for each dataset.

The Core Data Set Catalog prototype with direct links is based on the data of the former North Inlet (NIN) LTER site archived at LTERnet. Its entries are essentially the same as in the original Catalog, but dataset codes and metadata are hypertext links to files containing the complete archival information. If a Catalog search for *rainfall*, for example, leads to dataset NIN001 (Climate Data from North Inlet Meteorological Station with Water Parameters, 1982-1992), the datasets described therein can be retrieved by selecting the highlighted links. Similarly, complete metadata can be retrieved.

The prototype Image Catalog provides direct links to complete metadata (header files) and preview images (color, 512 x 512 pixel). Based on such considerations as cloud cover and extent of areal coverage, this information enables the user browsing the catalog to make an educated “pre-selection” of the image files desired. Given the several hundred megabytes per file, this capability can save considerable time and network resources.

4.0 Working Group Reports: Management of Spatial Data

4.1 Network of Networks

Leonard Walstad, Barbara Benson, Caroline Bledsoe, Marge Holland, Tom Kirchner, Kevin La Fleur, Linda May, Eda Melendez, Susan Stafford

This group discussed means by which communication among information managers could be enhanced, a sense of community for communication could be provided, and recommendations related to these. It was noted that pursuit of any or all of the following recommendations will require effort.

News groups: structure/hierarchy (by data types?) will need to be considered

Mailing lists: shared structure/categories with news groups; standardization of categories/headers/keywords/profiles would facilitate development of appropriate filters

EIM [Environmental Information Management] Spectator as a Mosaic Page; a periodical; editor(s); filtered form of news groups; concise format; identification of important information

Information Scout: specialist (“librarian”) in locating new information and structuring (filtering) that information for our community; could also serve as a focal point for information exchange between networks

There are already many (electronic) steps being taken in this realm. A human scout will quickly become overwhelmed with the increase in information; an electronic agent that automatically searches for relevant information might be preferable. A scout could establish an index and table of contents for Mosaic and help structure information. The need for *quality* Mosaic pages was emphasized.

4.2 Standards Development

Karen Baker, Gil Calabria, Scott Chapal, Harvey Chinn, David Fulker, Kenn Gardels, Merilyn Gentry, Bruce Gritton, John Helly, Rick Ingersoll, Clarence Lehman, George Lienkaemper, David Mark, Jim Quinn, Nancy Tosta

Note: A slide summary of this discussion, as prepared by Kenn Gardels, can be found at

http://www.regis.berkeley.edu/gardels/lter_standards.ps

In order to minimize redundant effort, the group decided to first determine the difference between this topic (i.e., standards development) and exchange of spatial data which was being addressed by a separate working group. It was concluded that standards development was broader in scope and that exchange should be more focused on the actual "plumbing".

A minimum set of standards must be practicable or they will not be implemented on a widespread basis. A description of standards should be technology-independent, i.e., information content requirements (and not technology) should "drive" standards development.

The group identified the key elements of metadata as 1) content, 2) representation, 3) structure, and 4) context. The first three are necessary to the "computer person" and the final element is required by the scientific domain user.

Extant standards (e.g., Ocean Data Evaluation System) should be examined. Descriptors such as instrument codes and collection codes facilitate transport, quality control/quality assurance, etc. It is critical that terminology be well-defined.

Although the paradigm of a hierarchy of standards (e.g., "top" level standards might not contain feature definitions) can be followed, the conceptual model of overall standards development must be a synthesis of all of the "layers" or "stacks" comprising the hierarchy.

At this point in the session, the group decided to direct its efforts toward describing the product (i.e., set of standards) and the means by which it can be generated. It also should be noted that the discussion appeared to become more focused on standards development within the context of the LTER network, despite the fact that LTER participants were in the minority.

Real observations must be distinguished from propagated data. The "bottom line" is that information content must be such that data can be used 100 years into the future. Documentation does not necessarily have to be limited to ASCII text. Examples were cited describing cases where scientists were able to improve the usability of a dataset by means of a videotaped interview with the original investigator. Methodology is absolutely essential; it was noted that percent vegetation cover data are useless without knowledge of the methodology. Video media could be used to document methodologies in many situations.

Information modelling was postulated as a means by which standards could be developed. This process can be simplified if no distinction is made between "data" and "metadata". A technology-independent "strawman" process was presented that consisted of the synthesis of external views (i.e., by the information manager, principal investigator, etc.). It would require that we 1) prioritize user classes, 2) identify the process to "run" the model, 3) validate the model against each external view, and 4) ensure that all data classes and information classes are covered. The system must be dynamic, i.e., evolvable.

Questions were raised (Can we “ride on the backs” of others? Are there starting points?). Initially, we must be guided by prioritization, of both users and the data that are likely to be shared. We must accept the fact that standards may only be achieved at the higher levels (at least for awhile) in some cases.

User views that should be included were 1) the site scientist (the one 10 years from now as well as the current one), 2) the visiting scientist, 3) the inter-site scientist, and 4) everyone else (e.g., general public, Congressmen).

The model could be applied to non-spatial (e.g., output from simulation models), as well as spatial data. Some laboratory data have no spatial component, other than where the laboratory is located. Moreover, not all data with a spatial component (e.g., from meteorological stations) are well-suited for GIS. Often documentation is not necessarily spatial in nature, e.g., “gray” literature, videotapes.

It was concluded that standards developed for the LTER sites as a whole, using the proposed model, could be achieved by (a) person(s) at the graduate student/post-doctoral level (perhaps between degrees) and possessing sufficient expertise and excellent communication skills. There was discussion concerning the person-years required (ranging from 1 to 5) that was not fully resolved when the working group adjourned.

4.3 Exchange of Spatial Data

Barbara Nolen, John Briggs, Mike Folk, Don Henshaw, Jeff Jefferson, Clarence Lehman, Bill Michener, Greg Shore, Cindy Veen

We determined that most spatial data exchange will involve the individual LTER sites, their state and other agencies operating at the same locations, LTER inter-site studies and other external groups interested in LTER spatial datasets. Aspects of spatial data exchange that need to be discussed include the legalities and liabilities of exchange. For instance, legalities due to licensing agreements on imagery purchased from EOSAT or SPOT or liabilities with regard to restrictions on endangered species or archeological site locations.

The Federal Geographic Data Committee has developed the Spatial Data Transfer Standards (SDTS). This is a possible vehicle for exchange. Each entity would need to be able to convert to and from SDTS for this to be useful. SDTS has developed a vector profile that has been tested and a raster profile that is untested. The LTER community could be helpful in testing these profiles and possibly implementing SDTS in spatial data exchange.

The format developed by SDTS would need to be investigated by the LTER community. Questions would include:

> How is metadata handled?
>
> Does SDTS conform to National Mapping Agency accuracies and standards?
>
> What about space efficiency, quick, convenient access, availability of being on-line in compressed format?

Spatial metadata was also addressed by this group and it is recommended that LTER develop a subset of metadata standards from the Content Standards for Digital Geospatial Metadata, such as PGBIO, which is a subset of metadata standards.

4.4 User Access and Image Catalogs

John Faundeen, James Frew, Mark Klopsch, Lolita Krievs, Robert MacArthur, Rudolf Nottrott

The group concluded that, for purposes of a catalog, an image is useless without metadata. Therefore, a “minimum” catalog entry would have to consist of metadata and some representation of the image. A catalog is by definition searchable. Based on metadata and image data searches, there are two types of search mechanisms that span a continuum.

1) Metadata: Full text (string) search vs. metadata fully maintained in a DBMS (allowing complicated queries)

2) Image data: A precomputed fixed “thumbnail” image vs. searches based on image content (e.g., greenness index or other statistical summaries).

The group concluded that it would be very useful and immediately feasible to start the process of image cataloging at the sites and proposed to create such catalogs using commonly available Internet access tools.

The access tool could be Mosaic (or another WWW browser), in which case the catalog entries would be encoded in hypertext markup language (HTML) on a site information server. As a minimum, a catalog entry should contain

Metadata for the image

A precomputed image in a browser readable format (GIF would probably be the format of choice)

A pointer to the source of the image. The image could be down-loadable if there are no proprietary or other restrictions. Otherwise the pointer would be to other information about availability.

Several sites indicated interest in implementing a simple image catalog with the above characteristics and will coordinate their efforts. A prototype of similar nature is available at LTERnet.edu for the remotely sensed LTER site images (Landsat and SPOT) archived at the Network Office.

4.5 Proprietary Issues

James Brunt

This group explored issues related to the proprietary nature of ecological data, within the context of developing on-line information systems. All agreed that new directions in ecological science and pressure from funding agencies would increase demand for on-line accessibility to data. Even so, there are still fears in the community stemming from these developments including those of misinterpretation, liability, being “scooped”, etc.

The group focused on means by which those fears could be allayed, as well as development of reward structures for improved data accessibility. Several suggestions, that could be implemented immediately, were made for quelling scientists' concerns. These were the development of data handling/sharing policies and inclusion, within the metadata for each dataset, an explicit use statement defining "domain" of the data; the domain would include the hypotheses and questions that the data were collected to address. For the future, it was clear that the "system" would have to provide a structure that encourages and rewards the open contribution of data. Such a structure would have to be implemented within/at several levels within the system, including peer groups, funding institutions, and tenure/promotion committees. The means by which these rewards could be generated were discussed, the most important of which was considered to be development and promotion of a system for review, publication, and citation of datasets identified as being particularly beneficial to the greater scientific community. Suggested mechanisms included (1) proposed guidelines for data citation to be provided to journal editors, (2) proposed guidelines for proposal review to be provided to funding agencies, and (3) the development of a data journal. [The latter mechanism was a working group topic during the LTER Business Session; see section 2.4.]

5.0 Working Group Reports: Inter-site Data Access

5.1 Sources of Funding for Database Development and Research Areas in Scientific Data Management

David Fulker, Scott Chapal, Jordan Hastings, Marge Holland, Richard Lent, Art McKee, Jim Quinn, Robb Turner, Leonard Walstad

Database-related funding is easy (per Susan Stafford's presentation), but then again, maybe not.

Ideas for worthwhile, doable database-related proposals must consider both institutional and technical matters. Technical research topics (leading to tangible products and useful to the entire community) discussed:

"Glue" to facilitate use of multiple tools

Coherent, quantitative descriptions of sampled data ---> good regional characterizations

Taxonomy of GIS data types ---> formal data model- --> software ---> data exchange, analysis and visualization

AI methods for QA/QC; also for classification

"Fuzzy" feature/boundary descriptions and their propagation through GIS and other analytic operations

"Institutional" research topics discussed:

Interagency mechanisms for coordinated entry/use of databases

Sociology of technology use and database access; ease-of-use/entry thresholds

Connections to GLOBE and other non-scientist networks (elderhostels...)

K-12 interactions

Responsibilities for keeping dictionaries, lists, registries...

Funding issues discussed:

Investigator-initiated or programmatic

Quick turnaround (small)

LTERnet ---> contract

Software support (beyond development/prototype)

Institutional topics ---> long-term funding

A suggestion was made to hold a workshop addressing funding mechanisms for environment-related DBMS efforts.

5.2 Metadata Standards and Exchange

Don Henshaw, Caroline Bledsoe, Gil Calabria, Harvey Chinn, Merilyn Gentry, Rick Ingersoll, Tom Kirchner, David Mark, Bill Michener, Greg Shore, Cindy Veen

The working group focused on the development of a standard for metadata content. It was recognized that coordination with the efforts of other organizations is essential. Similarities exist among environmental studies, and metadata will share similar content and terminology. The Future of Long-term Ecological Data (FLED) Committee is currently examining the existing criteria and standards used in documenting long-term ecological datasets. The development of a draft standard for LTER is in progress and will be shared with the FLED Committee when completed. The metadata discussion was limited to non-spatial data, as the efforts of the Federal Geographic Data Committee (FGDC) in developing standards for digital geospatial metadata have been thorough.

Metadata standards should incorporate both scientific aspects (e.g., rationale, methodologies), as well as computer science aspects (e.g., format). The idea of a minimum acceptable standard was discussed. What is the minimum set of information necessary when the user is not the originator of the data? Is the minimum set even greater for a remote user with no contact with the collection site? One expressed viewpoint was that one should not attempt to limit the metadata associated with a dataset in any exchange of data. Any minimum standard should be able to accommodate the maximum amount of metadata. Since the point of metadata is to make the numeric values useful to distant (in both space and time) researchers when the original investigators are not present to pass on an oral description of methods and measurements, the single standard we adopt should be the content necessary to ensure that usefulness. Another expressed viewpoint was that any “minimum” content standard will also be the de facto maximum content standard. Synthesis efforts will require complete sets of metadata.

How does one enforce any metadata standard? How does one involve the researchers in this process? A reward structure and incentives for properly documenting datasets needs to be established, perhaps through data publication. A peer-review system could enforce adequacy of metadata and could validate the metadata standards. It was noted that any outside review should always include people from the future. Scoring schemes could be established to rate the completeness and accuracy of metadata. The Spatial Data Transfer System (SDTS) has four levels of compliance to standards (0-4), where zero means no compliance, and level-one compliance is necessary for certification.

Action: The development of draft metadata standards will begin with the synthesis of every LTER site's metadata elements. This effort will capture all commonly used elements (attributes listed in most LTER sets of metadata) and other useful elements (those attributes listed in some LTER sets). Potentially, we could develop our own terminology, but it will be important to note synonyms with the FGDC metadata standard.

Once a set of metadata elements is agreed upon by the LTER data managers, the metadata set will be shared with the FLED Committee. This will allow FLED to synthesize the LTER standard along with other identified agency standards. At the next data managers meeting, we should plan to develop a consensus metadata standard based on the FLED Committee's recommendations. Currently, James Brunt is our liaison to the FLED effort, and future meetings should be coordinated with FLED representatives.

Data Exchange. Metadata and data will require standardized formats for exchange and for access by network searches. Standardized text formats such as netCDF are being used for these purposes. Most standardization efforts are being done internally with no wide-scale development over the scientific community. SDTS (Spatial Data Transfer System) has been developed to filter/convert data to and from GIS systems for exchange, but has not been widely used or tested.

Tom Kirchner of the CPR LTER site has developed a text-based common data interchange format for fixed field ASCII datasets. A server with conversion tools can convert datasets to and from common interchange format for transfer to other servers (which will require the writing of a new conversion unit for each external format). Ultimately, a SQL interface to the server would allow for query responses to be packaged in this common interchange format. The format could also be expanded to include binary or comma-delimited datasets.

Action: The Kirchner CPR format will be modified to match the LTER draft of minimum metadata standards. Kirchner will work with other LTER sites to develop metadata files in the CPR format, and to install the server at these sites for testing. The metadata standard along with this data interchange format may provide the basic ingredients for the planned LTER-wide information system.

5.3 Implementation of On-Line Information Systems Using Generic User Interfaces (e.g., Gopher, Mosaic)

Eda Melendez, Mike Folk, John Helly, Kevin La Fleur, Linda May

A primary goal of each LTER site must be to facilitate the sharing of data among scientists. If this were not one of our objectives, it would be impossible to justify our existence as a network and as a long-term program.

For this reason, one of the main concerns of the LTER data managers is to enable easy exchange of information within/across sites. This is not always easy, since the technology that makes this feasible is often at least one step “in front of us”. Many LTER sites already have implemented Gopher servers, but we need easier ways to access servers so that the data transfer can become a part of our daily routines.

This workshop was one of several that addressed some of these issues. We prepared the following list of resources that need to be established prior to connection to a Gopher and/or WAI server from their own workstations:

1. Infrastructure that make TCP/IP connections possible.
2. Platform configuration (e.g., a 4-MB RAM 386 computer, a 14.4 KB modem, and Trumpet Winsock. The same software will work on a workstation with Windows for Workgroups that is attached to an Ethernet LAN with Open Data-Link Interface (ODI) frame).

These components will facilitate access to a server and once they have been incorporated into the computational environment, data managers will need to familiarize themselves with the most efficient means of accessing remote data. All LTER sites should prepare their “home pages” for Mosaic. For the moment, we have guidelines prepared by the LTER Network Office, but we need other demonstrations and/or resources. The “NCSA Mosaic Web Index” includes pointers to a great deal of information on many important aspect of Mosaic and the Web:

http://turtle.ncsa.uiuc.edu/web-index.html

Mike Folk also provided a URL that contains “tutorials” available from NCSA. Tutorials that are currently available as Microsoft Word documents include:

Mosaic Tutorial

HTML Tutorial: Basic

Macintosh HTTP Server Tutorial

Other related tutorials in the works are:

HTML Tutorial: Advanced

IBM Compatible HTTP Server Tutorial

On-line versions of these tutorials are also under construction.

These documents can be found at the following URL:

http://www.ncsa.uiuc.edu/Edu/Tutorials/TutorialHome.html

5.4 Interfacing Generic Network Tools with Other Software (SQL Databases, etc.)

Rudolf Nottrott, Karen Baker, Barbara Benson, Darrell Blodgett, James Frew, Kenn Gardels, Mark Klopsch, Lolita Krievs, Kevin La Fleur, Robert MacArthur

Members of this group have found a number of tools on the Internet that can be useful in solving tasks commonly performed by data and information managers, and explored them to various degrees. These tools include:

- Spatial WAIS (URL **http://waisq.er.usgs.gov/wais/spatial.html**)
- Map viewers with selectable areas (**http://pubweb.parc.xerox.com/map**)
- Data browsers (pgbrowse/mibrowse, **ftp://ftp.sunet.se/pub/unix/databases/postgres/contrib**; also commercial browsers such as Oracle DataBrowse)
- Forms-based SQL query tools (GSQL **http://www.ncsa.uiuc.edu/SDG/People/jason/pub/gsql/starthere.html**)
- EOSDIS data distribution tools (version 0)
- Testbeds for Open Geodata Interoperability Specifications (OGIS)

The sources of these tools will be made available for further exploration on-line at LTER-net and in a hypertext version of the workshop proceedings.

They are potentially very useful to many of the working groups and projects (e.g., XROOTS, cf. in this report).

6.0 Outlines/Summaries for Invited Presentations

6.1 Data Management and Storage: Today and Tomorrow

James Frew (Institute for Computational Earth System Science, University of California, Santa Barbara)

- This is really 2 talks:
  - Technology trends affecting data management
  - The DBMS-centric data management paradigm

. . . wasn’t sure which one to give so you get them both...

- What I won’t talk about:
  - Data formats
  - GUIs
  - Object-oriented anything
  - ...

Technology Trends: An unrestrained, free-thinking session.

- Processor Technology Trends
  - Personal supercomputing
    - You already have a significant fraction of a CRAY on your desk
    - CRAY is shared, workstation is dedicated
    - CRAY real time?= workstation real time
  - Parallelism
    - new supercomputers: clusters of multiprocessors

        - only affordable way to build them
    - your lab: network(s) of workstations
        - more a software than a hardware change

- Processor Technology Example: Climate Modeling
    - NCAR Community Climate Model (CCM2)
        - publicly-available “benchmark” GCM
        - standard resolution: 144K cells (128 x 64 x 18)
            - CRAY Y-MP: 1.1 CPU hr / simulated yr
            - Alpha 3000/800: 200 CPU hr / simulated yr
        - coarse resolution: 34K cells (48 x 40 x 18)
            - Alpha 3000/800: 42 CPU hr / simulated yr
    - UCLA atmospheric GCM (AGCM)
        - coarse resolution: 14K cells (90 x 18 x 9)
            - (8) Alpha 3000/800: 24 CPU hr / simulated yr
- Processor Technology “Gotchas”
    - Programming massively parallel processors
        - fine-grained parallelism
    - Programming networks of workstations
        - coarse-grained parallelism
- Storage Technology Trends
    - Everything will be on- or near-line
    - Disk arrays will be preferred on-line storage
        - data layout (e.g., RAID) for performance and reliability
        - minimize physical impact (power, cables, interfaces)
    - “Tape arrays” will be preferred bulk storage
        - lots of small jukeboxes to maximize throughput
        - jukebox and media costs both scale up to terabytes
    - Optical storage is not cost-effective
        - only makes sense if you need random access to small things on archival media
        - otherwise, small media + expensive jukebox = big lose
        - price/performance static, vs. magnetic plummeting
- Storage Technology Example:
    - How to store a terabyte
        - Magnetic disk: $1000K
            - (500) 2 GB
            - 360 ... 2500 MB/s
            - media $1000/GB

- Optical disk: $500K
    - (1000) 1 GB + (6) 4-drive jukebox
    - 24 MB/s
    - media $100/GB
- Magnetic tape: $165K
    - (100) 10 GB + (10) 1-drive jukebox
    - 30 MB/s
    - media $ 3/GB

- Storage Technology “Gotchas”
    - Migration between on-line and near-line
        - different layout policies
            - disk: spread out to maximize transfer rate
            - tape: bunch up to minimize latency
        - different granularities
            - disk: file system
            - tape: “chunk”
    - How do you “back up” terabytes?
        - mirror writes
        - duplicate in background

Aside: *Why Most HSMs Don’t Work*

- HSM = hierarchical storage manager
    - “transparently” migrate files between disk and tape
- Latency doesn’t scale
    - disk/memory: 1000/1
    - tape/disk: 100,000/1: disk operations time out
- File granularity (“swapping”) doesn’t scale
    - typical disk object (small file) won’t fill memory cache
        - disk is random access: don’t need to fetch whole file
    - typical tape object (huge file) will fill disk cache
- Filesystem (e.g., NFS) is poor model
    - does work for small random-access (i.e. optical) files

- Network Technology Trends
    - WANs will be 20 ... 100 MB/sec
        - most scientific data will be transferred electronically
            - CDROMs will revert to “publication” status
        - real teleconferencing and telecollaboration

- ATM will take over the world
    - 155 Mb/s WAN; 155 Mb/s or 800 Mb/s LAN
        - first compatible LAN/WAN standard
    - pushed by phone companies
- ISDN is dead
    - only 128 Kb/s
        - V.34 modems already do 29 ... 115 Kb/s

- Network Technology Example: Electronic Data Delivery
    - ATM: 20 MB/s @ $10/GB

- 24-bit screen	4 MB	0.2 sec	$ 0.04
- CD-ROM	600 MB	30 sec	$ 6
- 8mm tape	5 GB	4.3 min	$50
- teleconference	100 MB	1 hr	$ 1

        - only needs 0.5 Mb/s
    - Getting faster and cheaper
        - 1997: 100 MB/s @ $1/GB
            - no point in mailing tapes unless it’s a truckload

- Network Technology “Gotchas”
    - Public utility politics
        - local monopoly may not provide connections
        - market pricing skewed by tariffs
        - Hollywood will dwarf science users on this one
    - Bandwidth availability
        - guaranteed bandwidth on a shared public WAN?
        - real-time protocols are a current research topic

- Technology Trend Summary
    - Design for the future
        - huge on-line archives (disk/tape “farms”)
        - everything on the network
        - massive parallelism
        - standard building blocks (processors, disks, ...)
    - Buy just-in-time
        - price/performance decreasing 40 . . . 60%/year
    - Watch out for non-scalable technologies
        - e.g., custom components, optical disks, ISDN, ...

DBMS-centrism: *A People’s Guide to Sequoia Thought*

- Earth Science: Data Management Goals
    - Manage data, not files
        - access by attributes and contents (vs. filename)
            - “ultraviolet solar irradiance in ice-free areas under the Antarctic ozone hole on or near 15 Oct 1991”
        - arbitrary granularity (vs. whole files)
        - track dataset lineage (vs. your memory ...)
    - Integrate data management and analysis
        - better storage/retrieval for existing analysis tools
            - systems: IDL, GRASS, S-plus, ...
            - languages: C, FORTRAN, ...
        - development environment for new analysis tools
- Data Management Must Support:
    - Existing datasets
        - most incoming data are already highly structured
    - Efficient loading
        - we want to load terabytes
    - Earth science data types
    - Data manipulation
        - keep the processing near the data
        - interact with external tools
- Data Management: Datasets
    - Dataset = data with a common source
        - e.g., satellite sensor, field campaign, simulation, ...
        - usually a collection of files in a common format
        - what most Earth scientists mean by “database”!
    - must maintain dataset integrity
        - much implicit, ill-defined metadata
            - “well known to those that know it well”
    - much overlap
        - common types, attributes, etc.
- Data Management: Efficient Loading
    - Support data in external files
        - avoids “copy in” overhead
        - allows direct access by existing tools
        - parallel loading: data->files, metadata->manager
    - Support multiple data representations

- minimizes reformatting, transformation, etc.
- maintains data integrity

- Data Management: Data Type Issues
  - Structure
    - e.g., array, arc, point, ...
  - Semantics
    - e.g., location, temperature, ...
  - Multiple representations
    - e.g., map projections, units of measure, ...
  - Shareability across datasets
  - Methods
    - e.g., map projection transformation
- Data Management: Processing
  - Eager vs. lazy
    - eager: generate “standard products” whenever inputs are available
    - lazy: don’t generate anything until explicitly requested
  - Local vs. remote
    - local: process data in/”near” the manager
      - I/O optimization
    - remote: process data wherever somebody can do it
      - speed/complexity optimization
  - These decisions must be dynamic
    - {eager,local} today; {lazy,remote} tomorrow ...
- The “DBMS-centric” Worldview
  - Science data management is end-to-end problem:

```
repeat {ingest
        store
        access
        retrieve
        analyze}
```

  - Database management system (DBMS) provides structure for end-to-end solution
- Earth Science Data Management: The “DBMS-centric” Solution
  - Ingest
    - log metadata
    - trigger “eager” processing
  - Store

- manage granules: files, fragments, blocks, (whatever ...)
- manage on-/near-line migration

- Access
  - via queries against metadata
  - via queries against content
  - via external filename
- Retrieve
  - seamlessly across/within storage granules
  - into direct client connection
  - into external file
  - trigger “lazy” processing
- Analyze
  - types/functions to manipulate Earth science data
  - able to orchestrate sequences of external programs

- A “DBMS-centric” Solution Using POSTGRES and Illustra
  - “Extended relational” DBMS
    - basic RDBMS functionality, derived from INGRES
    - “o****t-like” extensions: types, methods, etc.
  - Logical choice for Sequoia
    - developed by Sequoia co-director (Stonebraker)
    - POSTGRES free; with source code
    - Illustra is “commercial POSTGRES”; cheap for UC
  - Major emerging DBMS technology
    - where Gang of Four are headed
    - alternative to pure OODBMS
- POSTGRES/Illustra Extensions
  - User-specified types
    - multidimensional arrays, polygons, structures, ...
    - may be implemented as large objects
    - arrays can be “chunked” (tiled) for optimal access
  - User-specified functions
    - operate on native or user-specified types
    - written in POSTQUEL/SQL or C (dynamically loaded)
    - can specify function’s “cost” to aid query optimization
  - Large objects with file-oriented access
    - “External” files (i.e. outside DBMS)

        - Unix files owned by DBMS

    - POSTGRES experiments

        - "Inversion" files (POSTGRES only)

        - Direct access to tertiary storage

            - "storage manager" block-level interface

            - understands tertiary storage latency, volume issues

- Implications of DBMS Extensibility
    - Attribute-based file management

        - via metadata associated with large objects

    - Content-base file management

        - via DBMS functions accessing large object internals

    - Polymorphic types

        - methods convert between multiple representations

    - Indices on dynamic attributes

        - can build index on function return value

- Simple Example: Image Data Retrieval
    - A simple data retrieval problem:

        - "I want all AVHRR data you have for the rectangular region whose corners are Mt. Diablo and Santa Barbara, CA, for the month of May 1990."

    - ... expressed as a POSTQUEL query:

```
retrieve(result=clip(AVHRR,-122,38,-120,34))
   where month(AVHRR.date)=5
   and year(AVHRR.date)=1990
```

    - Returns names of newly-created images (clipped to fit search area)

- Complex Example: AVHRR Landcover Product

PostScript graphic available from author

- AVHRR Landcover Product: Processes
    - CAPTURE --> INGEST --> CALIBRATE --> COMPOSITE --> REGISTER --> PIXEL-MATH

        - names: processes outside the DBMS

        - arrows: large objects

- AVHRR Landcover Product: DBMS Support
    - Canonical function table

        - entry for each function (process box)

            - input/output parameter types

    - Lineage table

        -entry for each function invocation

            - --> function table entry

        - parameter values
    - large object processing history = name traceback
    - “alerter” (broadcast signal) issued on insert or update
- Distributed large object header (DLOBH)
    - large object “URL”: dlobh://server:port/database/...

- AVHRR Landcover Product:Middleware Support
    - Tcl “wrapper” between function and DBMS
        - passes parameters to function
        - passes lineage table entries back to DBMS
        - parse/assemble DLOBHs
        - use Tcl because
            - easy to code
            - command-level access to DBMS client library
    - Evaluation demons
        - one per host on which functions can run
        - lazy: run function when contacted by Tcl wrapper
        - eager: run function when receive lineage table alert
- DBMS-centrism Is Alive and Well
    - Landcover product is basis for Sequoia prototype of alternate EOSDIS architecture
        - Tcl middleware and DLOBHs will be “grease and glue” of Sequoia phase II computing environment
    - Image manipulation being expanded into full raster data support
        - SAIF schema
        - FGDC metadata
    - Ongoing work on coupling GCMs directly to POSTGRES
        - “The Big Lift”: GCM output streams directly into DBMS

6.2 GeoData Management Using Mapping, Imaging, and DBMS Software Tools

Kenn Gardels (University of California, Berkeley, and Open GIS Foundation)

Note: A slide summary of this presentation can be found at

http://www.regis.berkeley.edu/gardels/lter_geomodel.ps

The geomatics discipline has evolved in parallel tracks based on map-based resource inventory and cartographic production systems, image-based remote sensing and numerical modeling software, and relational and object-based database systems for managing spatial and non-spatial information. Each of these approaches uses a conceptual data model best suited to specific user requirements and system functionality. While these systems and data models have significant value to sectors of the geographic information community, it is apparent that hybrid systems and tools that can effectively utilize a broader array of geodata types are essential to effective information management and analysis. These provide greater geoprocessing flexibility, in a form more natural to users, while minimizing the need for complex geodata transformations and their associated information reduction. As network access to distributed information stores increases, it also becomes necessary to identify data modeling standards that ease data sharing. Current format-based methods are insufficient for dynamic, on-line query, and so client-server approaches and distributed object technologies for geodata and geoprocessing interoperability are rapidly bringing open systems solutions to geomatics.

Geodata Modeling: *An Overview of Geographic Information Management Using Mapping, Imaging and DBMS Software Tools.*

- Geomatics
    - geographic information technology - systems hardware, software, and data
    - software for capturing, storing, querying, analyzing, and displaying geographic information
    - common georeferencing framework - “real-world” coordinate system
    - range from toolbox of utilities to decision support system
- Geodata Paradigms
    - cartographic - layers of thematic information encoded as points, lines, areas, cells
    - georelational - attribute records (tuples) associated with location identifier
    - object oriented - features or entities comprising attributes of location and description
- System purposes
    - inventory and mapping
        - *production systems, with capabilities for digitizing (encoding map elements), editing, cartographic output, and library management*
    - query
        - *“what is where?” tools for identifying attributes of geographic entities, displaying specific phenomena, “desktop mapping”*
    - analysis
        - *procedures and algorithms for determining spatial coincidence/proximity, weighting and ranking thematic relationships, interpolating/extrapolating surface patterns, simulating temporal trends*
- Technical Requirements
    - coordinate systems/projections - conversion, rectification, affine transformation (rubber-sheeting); datum shifts
    - overlay - union, intersection, clipping, Boolean combination, logical comparison
    - spatial operations - selection, measurement, buffer, statistical summarization, aggregation/disaggregation, resampling, generalization

- Geodata Models
    - feature - simple or compound bounded entities - *objects*
    - region - continuous, distributed phenomena - *extents*
    - network - connected/directed nodes and segments - *graphs*
- Geodata Hierarchy

    See **http://www.regis.berkeley.edu/gardels/lter_geomodel.ps** for this diagram.
- Geographic Features
    - location based on description
    - assembled from spatial objects
        - single object
        - multiple objects with common geoposition
    - optionally aggregate to spatial datasets
    - incorporate thematic attributes
        - implementation of georelational model
        - association to data dictionary/catalog
    - include relevant metadata
- Spatial Objects
    - geometric objects
        - areas, lines, points
        - blocks, volumes
    - geographic reference
        - coordinate system
        - projection
    - topologic relationships
        - mathematical cell complexes
        - connection, adjacency, association
- Geographic Regions
    - description based on location
    - functions of spatial/temporal domain
        - cartographic coverages
        - image and photogrammetric scenes
        - observed/derived numeric fields
    - thematic content as dependent variable
    - can be viewed as spatial dataset
    - support metadata components
- Spatial Extents

  - discrete, stepwise
    - OID of polygons in tessellation
    - OID of segments in network graph
    - OID of points
  - continuous, sampled
    - digital elevation model
    - flow field samples
  - pixelated, averaged
    - digital imagery
    - raster cartography

- Geographic Networks
  - definition based on connectivity
  - explicit or relative geographic positioning
  - attributes on segments and nodes
  - behavioral characteristics for elements
    - direction/orientation
    - impedance, regulation, restriction
    - optimization
  - support dynamic/temporal/simulation modeling
- Spatial Graphs
  - geometric primitives - points, lines
  - topological relationships - connectivity, adjacency
  - geographic position secondary to spatial relationship
  - georeferenced nodes, abstract connectivity

Vector Structures

- irregular subdivision of space
  - disjoint
  - space-filling
  - overlapping
- geometric entities - Cartesian coordinates (*x*,*y*)
  - polygon - each entity fully described
  - arc-node - shared boundaries between entities
  - points & lines - degenerate case
- topology - spatial relationships
  - fundamental data structure - composition/decomposition of polygons, etc.
  - information content - explicit description of adjacency, connectivity

  - TIN - triangulated irregular network for surfaces (Delauney)
  - volumetric - extensions to faces, voids, solids, planes

- Raster Structures
  - regular subdivision of space
  - n-dimensional arrays - geographic location implicit in array offset (row/column)
  - valued at cells or pixels
  - indexing and compression
    - run-length encoding
    - quadtrees
    - tiles and chunks
  - multi-band and hyper-spectral imagery
  - hexagons, non-cartesian structures
- Point Structures
  - random or ordered sampling or identification
  - *x,y,z,t* positioning (explicit or implicit) plus attributes
  - valued at nodes (defined or arrayed)
  - basis for process modeling, finite element/difference analysis
- Geoprocessing
  - selection - feature extraction, display, measurement
  - map algebra - thematic overlay, comparison, numeric operations
  - cartographic modeling - spatiotemporal relationships, proximity analysis
  - network analysis - flow, pathfinding, dynamic segmentation
  - image processing - filtering, classification, band arithmetic
  - surface modeling - weighted distance, kriging, splines
  - geostatistics - distribution, correlation, clustering
- GIS Applications
  - automated mapping/facilities management
    - CAD model
    - vector graphics oriented
    - labels and identifiers
  - remote sensing/image processing
    - mathematical/statistical model
    - raster image oriented
    - extensions for “ancillary” (thematic or classified) data
  - spatial database management systems
    - feature or object based model

- record or tuple oriented
- space/geometry as attribute of object

- Hybrid GIS
  - georelational model linking geometry and attributes
  - raster/vector integration
  - toolbox approach
- Interoperability
  - requirement for access to multiple data models using diverse tools
  - evolution in technical approaches
    - special-purpose conversion tools
    - neutral data formats - comprehensive and coherent
    - distributed object technologies - messaging of objects encapsulating data and methods
  - schema merger strategies
    - shared data dictionaries
    - metadata catalogs and directories
    - feature definition mappings
- Open GIS
  - application of open systems concepts to geographic information management
    - shared data space based on generic, extensible, well-defined data model
    - interoperable applications that can pass data back and forth
    - heterogeneous resource browser for locating data and services on the net
  - APIs and accessible data models for COTS
  - standards, specifications, and protocols
    - SDTS, SAIF, DIGEST
    - SQL3/MM
    - Open Geodata Interoperability Specification (OGIS)
- OGIS Project
  - provide interoperability among multiple data models
  - allow network-based access to and utilization of distributed, heterogeneous data models
  - facilitate user/application access to any private data store
  - ensure metadata cataloguing capabilities for features and dictionaries
  - specify interfaces and services for any geospatial application
  - ensure industry consensus on specification
- OGIS Reference Model

  See **http://www.regis.berkeley.edu/gardels/lter_geomodel.ps** for this diagram.

- Prototypes
  - Sequoia 2000 - SAIF-based data store, PostGRASS/Guernewood Geoprocessor
  - CERES (California Environmental Resource Evaluation System) - OGIS-based tools for access to bioregional information centers
  - Alexandria - Digital library of spatial map/imagery information
  - *????? watch this space!!*

For more information...

- Guernewood
  - Guernewood plan and objectives

    ftp:01/07/95/s2k-ftp.cs.berkeley.edu/pub/sequoia/guernewood
  - PostGRASS documentation and code

    ftp:01/07/95/s2k-ftp.cs.berkeley.edu/pub/sequoia/guernewood/postgrass/src
  - GRASS 4.1 for DEC Alpha

    ftp:01/07/95/ftp.regis.berkeley.edu:/pub/grass/grass-axp-binaries.tar.Z
- OpenGIS
  - OpenGIS interoperability specification

    ftp:01/07/95/moon.cecer.army.mil/ogis/spec

    ftp:01/07/95/s2k-ftp.cs.berkeley.edu/pub/sequoia/schema/STANDARDS/OGIS

    http:01/07/95/www.regis.berkeley.edu/ogis.html

    http:01/07/95/moon.cecer.army.mil/ogis/discussion/ space.html
  - OGIS mail group

    send 1-line “*Subscribe <yourname>*” message to:

    ogis-request@moon.cecer.army.mil
- SAIF & SQL
  - Spatial Archive and Interchange Format

    ftp:01/07/95/s2k-ftp.cs.berkeley.edu/pub/sequoia/schema/STANDARDS/SAIF

    http:01/07/95/www.wimsey.com/~infosafe/saif/saifHome.html

    http:01/07/95/www.wimsey.com/~infosafe/saif/saif31spec.html

  SQL3

    ftp:01/07/95/speckle.ncsl.nist.gov/isowg3

    ftp:01/07/95/s2k-ftp.cs.berkeley.edu/pub/sequoia/schema/STANDARDS/SQL

6.3 The National Spatial Data Infrastructure

Nancy Tosta (Federal Geographic Data Committee)

The National Spatial Data Infrastructure (NSDI) is conceived to be an umbrella of policies, standards, and procedures under which organizations and technologies interact to foster more efficient use, management, and production of geospatial data. The Federal Geographic Data Committee (FGDC) is charged by the Office of Management and Budget and through Presidential Executive Order (EO #12906) with the responsibility of providing leadership in the development of the NSDI.

The FGDC is working in three primary areas to promote development of the NSDI. The first is establishing a clearinghouse through the use of metadata and the Internet to facilitate searching for and accessing geospatial data. This effort requires individuals and organizations to document their datasets, serve that documentation to the Internet, and use a variety of software tools, such as WAIS and Mosaic, on the Internet to search for and access data. The FGDC, with extensive input from the user community, developed and formally adopted (8 June 1994) a metadata standard to describe the content and characteristics of geospatial datasets to facilitate better management and the ability to share the data. The FGDC staff is developing training materials and conducting national, state, and regional workshops in the use of the metadata standard and Internet to promote this effort. Federal agencies are mandated in the Executive Order to begin to use the metadata standard in January 1995 for all new data collected and to make these metadata accessible electronically. By April of 1995 they are to have in place a plan for providing access to all geospatial data and are to be using the clearinghouse to search for data before expending funds for the collection of new geospatial data.

The second activity area is conceptualizing and testing the development of a digital framework dataset that will minimize redundancy in data collection and facilitate the integration and use of geospatial data. This dataset is envisioned to consist of the most commonly required datasets for most geospatial data applications, including digital orthoimagery, geodetic control, elevation, transportation, hydrology, administrative boundaries, and cadastral or ownership information. These themes of data will likely vary in resolution over different geographic areas and will likely be developed by different organizations, including state, local, and federal government agencies and the private sector. Over the last year, the FGDC Framework Working Group has developed a draft report on the framework concept which will be tested in a series of pilot studies in 1995. These pilots will help develop standards, identify institutional issues, and provide the foundation for programmatic changes and funding proposals for the framework.

The third activity area is the development of standards. The FGDC consists of a dozen subcommittees and numerous working groups that deal with a variety of themes of data, applications, and aspects of geospatial data management. The responsibility for leadership in the coordination of the various themes of data was assigned by the Office of Management and Budget to different federal agencies. Many of these groups are developing standards for data collection and content, data presentation, and data management to facilitate data sharing. For example, the Standards Working Group developed the metadata standard, the Cadastral Subcommittee is currently circulating a proposed standard for collection and representation of cadastral information, and the Wetlands Subcommittee is developing a consistent classification approach for mapping wetlands. All of the FGDC developed standards are subjected to an extensive public review process that includes nationally advertised comment and testing phases.

Development of the NSDI as a networked, distributed enterprise requires new relationships and partnerships among different levels of government and between public and private sector entities. The FGDC is promoting partnerships in a variety of ways. First, it is encouraging states to consider the formation of state and regional councils for coordinating geospatial data activities over the state or regional geography. These councils should consist of representatives of all of the sectors, agencies, and interests that collect or deal with geospatial data in that area. The FGDC will work with these councils to develop standards, facilitate and coordinate data collection efforts, and promote opportunities for data sharing. Second, the FGDC has established a Competitive Cooperative Agreements Program to encourage use and development of standards and the clearinghouse among nonfederal sectors. A maximum of $25,000 per project is available as seed money for creative partnerships to carry out these activities. Third, the FGDC is working on a database of current federal agency partnership opportunities related to geospatial data. Finally, the FGDC is developing the capability within the clearinghouse to identify entities who may be interested in data over the same piece of geography. This will facilitate the ability to find potential data development partners.

The NSDI is a continually evolving concept and process. Changes in GIS technologies, telecommunications, and institutions force adaptations in policies, procedures, and standards. Managing change is a constant challenge in evolving a robust NSDI.

6.4 Geospatial Data Acquisition, Quality, and Metadata for the Environmental Sciences

William K. Michener (Joseph W. Jones Ecological Research Center), David P. Lanter (Department of Geography, University of California, Santa Barbara, and Geographic Designs Inc.), and Paula F. Houhoulis (Joseph W. Jones Ecological Research Center)

Introduction. Geographic information systems (GIS) have evolved significantly as a technology that can promote understanding of environmental problems at local, regional, and global scales. As environmental scientists and managers increasingly direct their attention to broader scales, issues related to geospatial data acquisition and analysis will increase in complexity. Addressing questions on a regional level, for example, may entail acquisition of hundreds of geospatial datasets from dozens of different sources with variable accuracy, precision, and currency, as well as the use of up-to-date, high resolution satellite imagery.

Many of the constraints associated with utilizing GIS technology for environmental research and management have been removed during the past decade due to improvements in network technology, decreasing costs and increasing power of computer hardware and software, availability of a larger number of environmental geospatial databases, and new user interfaces that facilitate easy use of computer-driven applications. Concerns in the future will increasingly need to focus on how to best identify, acquire, manage, and analyze geospatial data that are timely, relevant to a specific problem, and of sufficient quality to support wise scientific and management decisions. In this report, we discuss three relevant issues and some potential technical solutions: data identification and acquisition, metadata (including source documentation and data lineage), and data quality documentation. Technology alone, however, will not entirely solve all relevant problems. The design and implementation of new GIS for supporting environmental research and management should incorporate metadata and lineage tracking throughout all project phases.

Identification and Acquisition of Relevant Data. Often we cannot find or access existing GIS databases because their contents and the meaning of those contents are unknown. One reason for this is that the meaning of data is exogenous and not found in the data (Tobler, 1979). In many situations, potential users are not aware that data they require already exist. Consequently, important studies are postponed until additional data are captured or created.

The *Geographic Information Explorer* (Geographic Designs Inc., 1994a) is a tool for visualizing, interacting with, and learning about the contents of environmental databases. The system facilitates understanding by providing users with abilities to interact with visualizations of the thematic, spatial, and temporal dimensions of a database's contents. Simultaneous interactive views of what data were collected, where they were collected, and when they were collected obviates which specific information requirements are met by available datasets and which ones remain to be filled.

The Geographic Information Explorer's design (Lanter and Essinger, 1994) is based on the principle that geographic facts can be well specified within three general dimensions: theme, space, and time (Berry, 1964; Sinton, 1977). Interactions with representations of how the data vary among these dimensions enables researches to:

- view the arrangement of locations within the context of a set of thematic attributes,
- view the set of thematic attributes available at a particular location,
- compare sets of thematic attributes as they vary across locations,
- compare locations in terms of the thematic attributes by which they are characterized,
- study a set of attributes at a set(s) of locations involving some or all of the above with the additional ability to study spatial association to enrich understanding of areal differentiation in terms of thematic characteristics of various sets of locations,
- compare locations through times in terms of the changing spatial distributions of collected thematic attributes,
- compare the set of thematic attributes collected at a location over time,
- compare changing spatial associations among thematic attributes,
- study changing areal differentiation (e.g., change in the nature of a habitat),
- compare a set(s) of locations and the associated thematic attributes over time, and examine the interplay of the preceding approaches.

These capabilities help meet the needs of researchers and analysts working to understand interrelationships among diverse geographically referenced data.

The Geographic Information Explorer treats each observation (i.e., data sample) stored within its object-oriented database as a composite assembled from spatial, thematic, and temporal components, each of which is treated as a complex object in its own right. A database is presented to the user with windows representing each of the three components. Thematic classes and associated attributes and values, locations, and time periods are each shaded to display the nature of the database that covers them. The result is a portrayal of the covariation of a database's contents.

User selections made in one dimension are immediately reflected in the other two dimensions. For example, a user selection of an attribute in the thematic window reflects in the other two, showing when and where it was measured. A user selection of a place, results in display of what was collected there and at which times. Selection of a time, displays what was collected then and at which locations. This enables the user to dynamically explore collections of datasets by manipulating theme, space, and time individually, or in concert as they:

- select data by specifying what is needed, where and when,
- explore covariation among different thematic variables, locations, and times,
- find places and times where data with chosen criteria coincide, and
- determine what additional data are available at the same time and place as a selected subset of data, within user-specified space and time constraints.

Metadata. Metadata describe the content, quality, condition, and other characteristics of data and represent higher level information about geospatial data that facilitate: (1) identification of available data for a particular geographic location, (2) determination of fitness for use of data for meeting a specific objective, (3) data acquisition, and (4) data processing and analysis (Federal Geographic Data Committee, 1994). There have been numerous attempts over the past decade at developing standards for spatial data transfer that incorporate a metadata component including: the *Spatial Data Transfer Standard* (National Institute of Standards and Technology, 1992), *Digital Geographic Exchange Standard* (Digital Geographic Information Working Group, 1991), and the *Vector Product Format* (Defense Mapping Agency, 1992). Recently, a comprehensive set of *Content Standards for Digital Geospatial Metadata* has been released (Federal Geographic Data Committee, 1994). The goal of physically integrating metadata and data within GIS has not yet been fully realized, although numerous vendors are devoting research and development funds to this activity, and the U.S. Geological Survey has developed DOCUMENT.AML, an Arc macro language program that facilitates manual documentation for Arc/Info coverages.

Despite the significant progress in the area of metadata standards, Chrisman (1994) asserts that all the standardized procedures in the world cannot ensure that the product actually satisfies the user's needs. He emphasizes the joint responsibilities of users and providers in relation to spatial data use and documentation, the need to incorporate spatial statistics more fully into GIS, research leading to a better understanding of error propagation in GIS and, importantly, the critical need to develop procedures that can handle large differences in resolution, accuracy and other key properties.

Data Lineage. Spatial data are often decentralized and partitioned across distributed networked computer systems. Most partitioning of data, however, is done randomly in response to project needs and available software and hardware. These fragmented data are often stored redundantly in different locations, on different databases, and in different data structures. They are typically named inconsistently, defined poorly, and documented incompletely. As a result, their meaning may have been lost and their quality unknown. They are generally unreliable, incompatible, incomplete, and inaccurate. They are difficult to understand, identify, and access, and expensive to maintain. As a result most databases contain massive quantities of disparate data that are poorly understood and cannot be reused (see Brackett, 1994).

If data within existing GIS databases are going to be of use to those outside their immediate users, they will have to be examined and defined element by element. This implies that each map within the GIS spatial database must be understood. The data must be analyzed to determine what the originators intended them to mean when they were initially created. This meaning, exogenous to the data itself, must be determined from available metadata. Lineage metadata (Lanter, 1994a) is a formal basis for determining the meaning of GIS derived spatial data. It consists of specific documentation pertaining to data sources, processing history, and product use (Lanter, 1991). The National Committee on Digital Cartographic Data Standards' Spatial Data Transfer Standard (SDTS, 1992), recommends that source documentation include name, feature content, dates, responsible agency, scale, projection, and accuracy elements.

Lineage metadata provides a basis for uniquely identifying, and defining data and specifying how GIS derived maps relate to other relevant maps in the database. Automation of lineage metadata makes it possible to optimize the contents of the spatial database (Lanter, 1993), model errors propagating within spatial analyses (Lanter and Veregin, 1992), identify and remove redundancy and propagate updates through a distributed spatial database (Lanter, 1994a), and generalize spatial analytic logic used within GIS applications (Lanter, 1994b).

Geolineus (Lanter, 1992; Geographic Designs Inc., 1994b) automates lineage metadata to support GIS users in understanding, documenting, updating and managing (GIS) databases. Lanter and Surbey (1994) pioneered the use of Geolineus as a data discovery tool for assessing the quality of disparate data derived within prior GIS applications databases. As Lanter and Surbey's metadata analysis illustrated, reverse engineering data quality documentation requires thought, analysis, intuition, and consensus by knowledgeable people to identify the true content and meaning of disparate data. An automated tool, such as Geolineus, can support people in the data engineering process, but it cannot create missing documentation or clearly define unknown prior data.

Dealing With Uncertainty. Increased use of GIS for research and management, and the creation of multiple use, widely-shared geographic databases requires that data quality and the data entry process be closely examined and well documented (Dangermond, 1988). Spatial data are prone to uncertainty and inaccuracy and the reliability of any computer-based analysis will be constrained by data quality (Goodchild et al., 1993). There is a natural tendency to ignore the fact that maps are only approximations of the truth, representing subjective interpretations of real spatial variability. Burrough (1986), for example, asserts that many soil scientists and geographers know from field experience that carefully drawn boundaries and contour lines on maps are elegant misrepresentations of changes that are often gradual, vague or fuzzy. People have been so conditioned to seeing the variation of the Earth's surface portrayed either by the stepped functions of choropleth maps or by smoothly varying mathematical surfaces that they find it difficult to conceive that reality is otherwise. Consequently, the ease of access and manipulation provided by GIS, and its inherent precision, encourages users to see spatial data as accurate, and to lose touch with the uncertainties of the measurement and capture process (Goodchild and Gopal, 1989; Goodchild et al., 1993).

Many of the environmental geospatial datasets currently available in digital form (e.g., soils, wetland inventory data) were originally derived from digitization of cartographic maps based on interpreted aerial photography. Thus, these data are subject to various primary and secondary data collection errors (Thapa and Bossler, 1992). Primary data collection errors include personal, instrument, and environmental errors. Secondary data collection errors are attributable to numerous aspects of cartographic and GIS processing, including: plotting control points, compilation error, drawing or plotting, map generalization, map reproduction, color registration, material deformation (humidity changes, etc.), changing scales, uncertainty in feature definition, feature exaggeration, digitization or scanning, and labeling and feature coding.

Cartographic standards have been relatively stable for several decades. For example, if 90% of well-defined points fall within 0.5 mm of their true position, then the map can be certified as complying with the US National Map Accuracy Standard (NMAS) (Bureau of the Budget, 1947). Although very effective for roads, utilities, and other easily defined features, the NMAS is not as easily translated to many environmental features that are, by nature, fuzzy and non-linear. For example, the reliability of soil series maps can range from 10% to 82% (see review by Fisher, 1989). When soil inclusions were represented as distinct soil mapping units, the overall accuracy of the USDA Soil Conservation Service soils maps was assessed at approximately 60% (Walsh et al., 1987).

Additional errors may be incorporated into GIS analyses when data layers are based on processed satellite imagery. For example, although Jensen et al. (1993) were able to classify coastal wetland and upland habitats with 86 to 92 % overall accuracy, efforts to differentiate among some classes (bare soil, cultivated land, herbaceous) and among the various wetland types (estuarine, riverine, marine) were not feasible. Errors may be further compounded when satellite data at the pixel scale are aggregated into larger units. For example, landcover classification accuracies ranged from 57% and 43% for multispectral scanner data aggregated to 100 and 200 m^2, respectively, and errors in the categorization of slope angle and aspect were consistently over 50% for digital elevation models summarized at 100 and 200 m^2 (Walsh et al., 1987).

Various GIS procedures lead to non-linear and often counter-intuitive propagation of errors. For example, addition of data layers of fixed accuracy results in a negative exponential error rate (Veregin, 1989), whereas the buffer operation can actually lead to increases in overall accuracy for the final product (Veregin, 1994). Finally, GIS may be leading to uses for which the data were never intended when they were collected and mapped (Goodchild, 1993). It is important, therefore, to have well-documented metadata and known sources or contacts, in order to determine fitness for use.

Conclusions and Recommendations. Many of the constraints associated with utilizing GIS technology for environmental research and management have been removed during the past decade. However, despite numerous database and technological improvements, several impediments remain. First, how do we identify and acquire data of appropriate scale (spatial, temporal, thematic) and quality required for a particular project? Secondly, how can we be assured that the cartographic products resulting from a series of GIS and modeling operations are both geographically/statistically significant and environmentally meaningful?

Technology may ameliorate the problems, but will not entirely resolve them. Ultimately, GIS design and implementation should reflect the importance of data and metadata-based data quality assessments, and metadata development and lineage tracking throughout all phases of research and resource management projects. New software tools, such as Geographical Information Explorer, can facilitate identification of appropriate data. Similarly, Geolineus, can automate all or portions of the process of tracking data lineage throughout data management, GIS, and modeling operations.

Ideally, errors associated with each of the data layers would be propagated through the series of analyses to evaluate uncertainty associated with the final product(s). The process of tracking error within GIS has been automated (Lanter and Veregin, 1992). Although much basic research remains to be done in this area, new understanding is emerging regarding the manner in which errors are affected by some specific GIS operations (e.g., AND/OR operations [Veregin, 1989; Lanter and Veregin, 1992] and buffering [Veregin, 1994]). Thus, it is currently possible to overlay coverages of known accuracy and examine the magnitude and range of errors that may result in final products. Such analyses can prove critical for determining whether available geospatial data are adequate for meeting project objectives or what level of accuracy is required in new data layers.

GIS-based environmental research and resource management depend upon the quality of available data. If *a priori* consideration is paid to metadata and data quality issues, then organizations can focus valuable time and effort on performing appropriate analyses with the requisite high quality data. As data lineage tracking and metadata maintenance programs are formally incorporated into project planning and implementation, organizations can further benefit by being able to re-use data developed for other applications.

Acknowledgments. Arc/Info is a trademark of Environmental Systems Research Institute, Inc.; Geolineus is a trademark of Geographic Designs, Inc.

Literature Cited:

Berry, B. 1964. Approaches to regional geography: A synthesis. Annals of the American Association of Geographers 54: 2-11.

Brackett, M.H. 1994. Data Sharing - Using a Common Data Architecture. New York: John Wiley & Sons, Inc.

Bureau of the Budget. 1947. National Map Accuracy Standards. Washington, D.C.: U.S. Government Printing Office.

Burrough, P. A. 1986. Principles of Geographic Information Systems for Land Resource Assessment. Oxford: Clarendon.

Chrisman, N. R. 1994. Metadata required to determine the fitness of spatial data for use in environmental analysis. Pages 177-190 IN W. K. Michener, J. W. Brunt, and S. G. Stafford, editors. Environmental Information Management and Analysis: Ecosystem to Global Scales. London: Taylor & Francis.

Dangermond, J. 1988. GIS trends and comments. ARC News. Summer/Fall Issue, pp. 13- 17.

Defense Mapping Agency (DMA). 1992. Vector Product Format, Military Standard 600006. Washington, D.C.: Department of Defense.

Digital Geographic Information Working Group (DGIWG). 1991. DIGEST: A Digital Geographic Exchange Standard. Washington, D.C.: Defense Mapping Agency.

Federal Geographic Data Committee (FGDC). 1994. Content standards for digital geospatial metadata (June 8). Washington, D.C.: Federal Geographic Data Committee.

Fisher, P. F. 1989. Knowledge-based approaches to determining and correcting areas of unreliability in geographic databases. Pages 45-54 IN Goodchild, M. and S. Gopal, editors. The Accuracy of Spatial Databases. New York: Taylor & Francis.

Geographic Designs Inc. 1994a. Geographic Information Explorer Version 2.02. Santa Barbara, California, USA.

Geographic Designs Inc. 1994b. Geolineus 3.0 User Manual. Santa Barbara, California, USA.

Goodchild, M. 1993. Data models and data quality: Problems and prospects. Pages 94- 103 IN Goodchild, M. F., B. O. Parks, and L. T. Stayaert, editors. 1993. Environmental Modeling with GIS. New York: Oxford.

Goodchild, M. and S. Gopal. 1989. Accuracy of Spatial Databases. New York: Taylor & Francis, 290 pp.

Goodchild, M. F., B. O. Parks, and L. T. Stayaert. 1993. Environmental Modeling with GIS. New York: Oxford, preface.

Jensen, J. R., D. J. Cowen, J. D. Althausen, S. Narumalani, and O. Weatherbee. 1993. An evaluation of the CoastWatch change detection protocol in South Carolina. Photogrammetric Engineering and Remote Sensing 59(6): 1039-1046.

Lanter, D. P. 1991. Design of a Lineage-Based Meta-Database for GIS. Cartography and Geographic Information Systems 18(4): 255-261.

Lanter, D.P. 1992. GEOLINEUS: data management and flowcharting for ARC/INFO. Technical Software Series S-92-2, National Center for Geographic Information and Analysis, Santa Barbara, California, USA.

Lanter, D. P. 1993. A lineage meta-database approach towards spatial analytic database optimization. Cartography and Geographic Information Systems 20(2): 112-121.

Lanter, D. P. 1994a. A lineage metadata approach to removing redundancy and propagating updates in a GIS database. Cartography and Geographic Information Systems 21(2): 91-98.

Lanter, D. P. 1994b. Comparison of spatial analytic applications of GIS. Pages 413-425 IN Michener, W.K., J. W. Brunt, and S. G. Stafford, editors. Environmental Information Management and Analysis: Ecosystem to Global Scales. London: Taylor & Francis.

Lanter, D. P. and Essinger. 1994. The Environmental Data Explorer, An intelligent interface for exploring unfamiliar environmental data sets. Geographic Designs Inc., Santa Barbara, California, USA.

Lanter, D.P. and C. Surbey. 1994. Metadata analysis of GIS data processing: A case study. Pages 314-324 IN Waugh, T.C., and R.G. Healey, editors. Advances in GIS Research. Proceedings of the Sixth International Symposium on Spatial Data Handling. London: Taylor & Francis.

Lanter, D. P. and H. Veregin. 1992. A research paradigm for propagating error in layer- based GIS. Photogrammetric Engineering and Remote Sensing 58(6): 825-833.

National Institute of Standards and Technology. 1992. Spatial Data Transfer Standard. Federal Information Processing Standard 173. Gaithersburg, Maryland: National Institute of Standards and Technology.

SDTS. 1992. ASTM Section D18.01.05 Draft Specification for Meta-Data Support in Geographic Information Systems. Information Exchange Forum on Spatial Metadata. Federal Geographic Data Committee, U.S. Geological Survey, Reston, Virginia, USA.

Sinton, D. 1977. The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study. Proceedings of an Advanced Study Symposium on Topological Data Structures for GIS, Volume 1. Harvard Department of Landscape Architecture, Cambridge.

Thapa, K. and J. Bossler. 1992. Accuracy of spatial data used in geographic information systems. Photogrammetric Engineering and Remote Sensing 58(6): 835-841.

Tobler, W.R. 1979. A Transformational View of Cartography. The American Cartographer 6(2): 101-106.

Veregin, H. 1989. Error modeling for the map overlay operation. Pages 3-18 IN Goodchild, M. and S. Gopal, editors. The Accuracy of Spatial Databases. New York: Taylor & Francis.

Veregin, H. 1994. Integration of simulation modeling and error propagation for the buffer operation in GIS. Photogrammetric Engineering and Remote Sensing 60(4): 427- 435.

Walsh, S. J., D. R. Lightfoot, and D. R. Butler. 1987. Recognition and assessment of error in geographic information systems. Photogrammetric Engineering and Remote Sensing 53(10): 1423-1430.

6.5 Models, Data Formats, and Software Tools Applicable to Inter-site Data Exchange

David Fulker (Director, Unidata Program Center, University Corporation for Atmospheric Research)

Introduction. The purpose of the Unidata Program Center (UPC) is to provide software and services that empower universities to acquire and use atmospheric and related data on their own computers, often in real time. Funded by the National Science Foundation and managed by the University Corporation for Atmospheric Research, the UPC serves more than 120 universities by: facilitating access to real-time data, providing specialized software systems, and fostering participation in Unidata as a community endeavor. Software-related activities at the UPC encompass development, upgrades, maintenance, training, consultation, and documentation.

Software provided by Unidata falls into two categories: 1) analysis and display applications and 2) data management. The analysis and display applications have a distinct meteorological emphasis and, in general, they have been developed in the community, not at the UPC -- the Unidata role is to provide software support. In contrast, data management tools and systems are both developed and supported by Unidata staff. Their purpose is to provide underpinnings for (generic) analysis and display applications, and there is strong emphasis on distributed functionality. Because of this emphasis, Unidata is able to meet community needs without a data center, a rather nontraditional approach.

Data Management in Unidata. For this audience I will emphasize data management rather than Unidata’s (meteorologically oriented) analysis and display capabilities. Within the data management arena, Unidata offers two primary software packages. The first, dubbed the Local Data Manager or LDM system, supports real-time reception of multiple data streams to created tailored (i.e., site-specific) data holdings and to manage event-driven processing of data immediately upon arrival. The LDM employs a client-server architecture that allows LDM systems to handle intra-site and well as inter-site data flows, exploiting the power of the Internet.

The second major Unidata package in the data management arena is the Network Common Data Form (netCDF) library. This highly portable software system functions somewhat like an I/O library, providing means to store and retrieve -- by name and index -- multidimensional arrays and ancillary data through Applications Programming Interfaces (APIs) for both FORTRAN and C. The underlying format that defines how the data are actually stored in a file is highly portable.

This leads to the primary focus of this talk: effective use of data models, data formats, and software tools to facilitate inter-site data exchange. I will attempt to define the term “data model” and provide examples. I will touch on data formats by mentioning some pitfalls of “standard formats” and describe the benefits of data hiding as an alternative. I will visit as well the subject of software tools for managing data, noting that there are important issues surrounding portability, APIs, and efficiency. Finally, I will illustrate these concepts by describing the netCDF library in some detail.

Data Models. A data model may be defined as an abstraction of a dataset or a database, one which characterizes the types of data represented and defines certain relationships among the data elements. The model also specifies the available operations, such as definition, storage, modification, query, and retrieval.

There are many examples of data models. Within the most common programming languages are the models which underlie FORTRAN read and write statements or the standard *i/o* and *scanf* functions in C. Other models are implemented via separate libraries. Examples include HDF and netCDF. The model most widely known and understood is the relational model underlying many database management systems and their Structured Query Languages (SQLs).

Data Formats. One may consider a data format to be part of a particular data model. In such a model all data are represented as sequences of bits on a storage "device," such as a disk file or a tape. Each instance of a format applies to one or more sequences of bits and (combined with knowledge of the machine architecture in use) defines a precise mapping between these sequences and a collection of data values whose types may encompass floating point numbers, character strings, and the like. Thus a format serves to identify data types and to define sequential relationships among data elements.

The context in which data formats are employed typically supports a rather small collection of read/write operations. They may be strictly sequential (where data must be read in the same order as they were written) or they may be direct (where an offset, usually a count of octets or bytes, defines the point in a stored sequence at which reading or writing is to begin). Each read or write transfers a sequence of bits, and these are converted to or from other data types under the control of a format.

Inter-site data exchange requires not only that sequences of bits be transported from one location to another (via network or portable media) but also that the data provider and data recipient employ the same formats for a given bit sequence, possibly compensating as well for differences in machine architectures. These problems give rise to a "standard" format, the purpose of which is to give providers and recipients all of the information necessary to exchange data successfully while requiring them to employ only a small number of data formats (perhaps one).

Pitfalls of standard formats. Unfortunately, standard formats are of limited value in facilitating inter-site data exchange. It is difficult to ensure compliance, i.e., to be certain that provider and recipient are, in fact, using precisely the same format, especially if their machine architectures differ. Even with a standard format, significant programming effort can be required for each context in which the data are to be used. And because the kinds of data that must be exchanged are rarely static, a "standard" format often must be changed to accommodate new data or changes in the organization of data.

In summary, standard formats are prone to errors and misuse, and often they are too static. One approach to solving this problem is known as data hiding.

Data hiding. By using a higher level of abstraction, users need not know the format to exchange data effectively and correctly. In this approach, one develops a data model that encompasses all of the data types, relations, and operations that are needed by providers and recipients to represent and use the information to be exchanged. Then one develops a software system which can be used on the computers of providers and recipients, whose Applications Programming Interface (API) realizes all aspects of the model, and which employs data formats chosen to achieve the desired level of portability. Finally, providers and recipient perform *all* data access (i.e., storage and retrieval) through the API.

If such software is well written, it enforces compliance with the underlying format, reduces the likelihood of errors, compensates for architectural differences, and reduces programming effort because formats never appear in user codes. Finally, if the data model is sufficiently general, new data or changes in the organization of data can be accommodated without modifying the API and, therefore, without modifying users' codes -- a significant improvement upon circumstances in which small modifications in a format require all users to modify their programs.

Software Issues and Decisions. Of course, creating a data model, designing and realizing an API, and defining a suitable underlying format for inter-site data exchange are not simple matters and require addressing at least the following issues. First, one must consider the support required for users of the data; the tools needed by a very small community of sophisticated users may be quite different than those that would be suitable for a large group of relatively naive users.

One must also consider the computing environments in which the data are to be used; achieving portability -- both of software and of data -- across a set of Unix-based platforms is simpler than doing so across a broader set of computers and operating systems. The data model and the underlying format must be file-system compatible and consistent with the computer languages employed by providers and recipients of the data. For example, if inter-site data exchange is to occur via the Internet file transfer protocol (FTP), this imposes certain constraints on how a dataset can be represented.

Finally, matters of efficiency must be considered, both in the context of data storage and data access operations. Inevitably, "trade-offs" must be made among conflicting factors such as compactness, portability, access to subsets, and simplicity of the API, and these can be decided only in the context of a given community of users.

Unidata's netCDF Library. I conclude by describing the software library developed at the UPC to facilitate scientific data access and inter-site data exchange. Our target community comprises geophysical scientists, primarily those in the atmospheric and oceanic disciplines, and it is relatively large; we provide (on-line) software support for hundreds of users. In our context, portability of data and software encompasses all of the most common computers and includes supercomputers.

The netCDF library creates and reads datasets that are represented -- in their entirety -- as FTP-transportable binary files, and the method is compatible with both the FORTRAN and C programming languages. Specifically, there are FORTRAN and C APIs; data stored by a FORTRAN program, for example, can be FTP'd to a dissimilar computer and retrieved in a C program.

In choosing among efficiency trade-offs we chose a highly portable and reasonably compact (but uncompressed) data representation based on Sun's eXternal Data Representation (XDR). This imposes a small performance penalty on access because all data pass through an XDR software layer.

The netCDF library realizes a direct access model for multidimensional scientific data. The interface (i.e., the C and FORTRAN APIs) insulate applications from data representation; that is, they “hide” data as discussed above. In addition, the library is especially well suited to inter-site data exchange on a large scale, because the software is highly portable and netCDF datasets are self describing and platform independent.

netCDF: the model. The three primary elements of a netCDF dataset are dimensions, variables, and attributes. Each dimension has a name (e.g., time, latitude, longitude) and a size, and sets of dimensions are used to specify the shapes of variables and define coordinate systems. Variables also are named (e.g., temperature, pressure, time) and can be scalars, vectors, or multidimensional arrays of type real, integer, or character. Named attributes (e.g., units) carry information about variables or the whole dataset.

Used together, variables, dimensions, and attributes can help capture the meaning of data. Specifically:

- a dataset has dimensions, variables, and attributes
- a dimension has a name and size (possibly unlimited)
- a variable has a name, type, shape, values, and attributes
- an attribute has a name, type, and value(s)
- a variable’s shape is specified by a list of dimensions
- a dimension may have an associated coordinate variable

The following example may help make the abstraction concrete. It describes (without actual data) a netCDF dataset by employing a simple notation we have defined as the common description language or CDL.

- Dimensions

• lat	73
• lon	73
• time	unlimited

- Variables

• float	P (time, lat, lon)
• float	lat (lat), lon (lon), time (time)

- Attributes

• P:long_name	“mean sea level pressure”
• P:units	“hectopascals”
• P:valid_range	800., 1200.
• lat:units	“degrees_north”
• lon:units	“degrees_east”
• time:units	“hours”

netCDF: the interface. The netCDF API is designed to be general enough for a broad range of programs that read and write scientific data. In the netCDF interface, data points are not accessed sequentially by individual reads and writes, but rather by specifying a variable and indices which determine the portion of the variable to read or write. The data types most commonly needed for scientific data -- bytes and characters, short and long integers, single and double precision real numbers -- are supported. Something we call *hyperslab access* permits reading and writing data cross-sections with a single function call.

One unlimited dimension is allowed, and new data may be appended to an existing dataset along that dimension. Thus, the unlimited dimension is like a record number in conventional record-oriented files.

Example netCDF calls in C:

- Open a netCDF file for read-only access --
  ```
  ncid = ncopen (“mydata.nc”, NC_NOWRITE);
  ```
- Get the variable ID for a variable named “velocity” --
  ```
  vid = ncvarid (ncid, “velocity”);
  ```
- Get *n*th variable’s name, type, shape, and number of attributes --
  ```
  ncvarinq (ncid, n, name, &type, &ndims, dims, &natts);
  ```
- Read a slab of values --
  ```
  ncvarget (ncid, varid, start, count, vals);
  ```

Example netCDF calls in FORTRAN:

- Open a netCDF file for read-only access --
  ```
  NCID = NCOPN (‘data.nc’, NCNOWRIT, IERR)
  ```
- Get the variable ID for a variable named “velocity” --
  ```
  VID = NCVID (NCID, ‘velocity’, IERR)
  ```
- Get *n*th variable’s name, type, shape, and number of attributes --
  ```
  CALL NCVINQ (NCID, N, VNAME, VTYPE, VRANK, VDIMS, VATTS, ERR)
  ```
- Read a slab of values --
  ```
  CALL NCVGT (NCID, VID, START, COUNT, VALS, IERR )
  ```

There are differences: under the C interface, indices start at 0, and the left-most dimension varies fastest; under the FORTRAN interface, indices start at 1 and the right-most dimension varies fastest. These affect only the programming interfaces; data storage is equally efficient through either language.

netCDF: the software. Both source and binary versions of the netCDF software library are freely available via anonymous FTP from --

host:	**unidata.ucar.edu**
file:	**pub/netcdf/netcdf.tar.Z**

The software has been ported to many platforms, including --

- Unix-based:
  Sun SPARCstation, IBM RS6000, DEC Alpha, DEC Ultrix, CRAY YMP, SGI Iris, HP 9000, ...
- VMS-based:
  DEC VAX, DEC Alpha
- Others:
  MSDOS and OS/2 for PCs

The software distribution includes comprehensive test routines and test data for verifying successful installation.

netCDF: the format. As mentioned above, the format for netCDF files employs XDR, a non-proprietary standard software layer for describing and encoding data in a machine-independent way (using IEEE standards where appropriate). XDR has been implemented on most computers from PCs and workstations to mainframes and supercomputers. On many systems the XDR representations, for integers, characters, floating-point numbers, etc., are identical to the native representations for those entities; on such computers absolutely no precision is lost when data are stored in netCDF files.

It is unnecessary to know the format to write and read netCDF data; successful software installation is all that is required.

netCDF-compatible software. Provided with the netCDF software library are two utilities: *ncdump* and *ncgen*, which provide means for creating textual descriptions of netCDF datasets (in the common description language, CDL) or, alternatively, creating netCDF files from textual descriptions. These tools are quite powerful, and the textual descriptions for netCDF datasets often are used in human discussions about how data are to be organized and represented in files.

Below is a short list of applications that read or write netCDF files.

- IVE (University of Washington)
- SpyGlass Dicer (Spyglass, Inc.)
- IDL (Research Systems, Inc.)
- WXP (Purdue and Unidata)
- SIEVE (USGS)
- HDF/netCDF merger project (NCSA)

For a longer list, see the netCDF entries in Unidata’s World Wide Web server at

http://www.unidata.ucar.edu

That server provides additional information about Unidata in addition to greater detail--including a comprehensive Users Guide -- on the netCDF software library.

The netCDF may or may not be well suited to the specific needs of the LTER community -- my intention is not to convince you to become users. Rather, I hope our experience with the netCDF, and the success the package has achieved, serve to illustrate the value of approaching the inter-site data exchange problem from the perspective of data models and software tools that hide data formats.

6.6 Integrating Modern User Interfaces and Scientific Data Formats

Mike Folk (National Center for Supercomputing Applications, University of Illinois (Urbana/Champaign))

The following is an outline/abstract of the presentation. It does not contain the overhead transparencies, nor the material that was displayed in the demonstrations. An illustrated version of the talk can be found at the following URL:

http://hdf.ncsa.uiuc.edu:8001/presentations/LTER/talk.html

What do we mean by “modern user interfaces”? Modern user interfaces can mean several different things:

- Clients: Gopher, Mosaic, etc.
  GUI-based applications
- GIS applications
  Dataflow-programming systems like AVS
    - NCSA Collage
    - GEOLineus
- APIs
  IDL, HDF, netCDF, etc.
    - These imply specific data models.

  Scientific data servers

Standard data formats? By standard data formats we mean formats like HDF, netCDF, SDTS, Digest, etc.

An evolutionary view. Over the past decade we have evolved from the “snail mail” model of access to one based on developments that are currently evolving, including such things as client/server technologies.

- Snail mail model
  There were no wide-area networks.

  Virtually all data had to be transmitted by mail.
- FTP model
  Leverage off of the Internet

  Transfer entire files
- Work in progress
  Client-server technologies

  WAIS, Gopher, WWW, etc.

  Collaborative technologies

Data management technologies

Information management technologies

leading to . . .

- Client/server model

  Leverage off of workstation technologies on client-side

  Leverage off of server-side technologies like data and information management

Client/server examples, focusing on WWW and Mosaic.

Client technologies = browsers and servers

Server technologies = data and info management, and scalable server work

- Client-side examples

We demonstrated the X-Window version of NCSA Mosaic viewing HDF and netCDF files. **Mosaic** . . .

Creates an HTML synopsis of the contents of an HDF or netCDF file

Displays in-line GIF images created from raster images (here we used Mosaic to view harsh.hdf)

Displays overview of file contents (here we viewed sag1.ozn...using Mosaic)

Allows you to look more deeply, at attributes (here we viewed test.hdf using Mosaic (sag1.ozn...)

Allows you to look at associated info, via annotation with hyperlinks (viewed test.hdf, then clicked on a hyperlink in an annotation)

We also looked at a netCDF example (z-2d.nc).

Mosaic shown talking to another application (**Collage**):

- Can view dataset in Mosaic then send it to Collage (viewed sstday.hdf in Mosaic, then clicked on “here” to send it to Collage. Note: Collage must already be running.)

- Sends data to Collage for further viewing

- In future, CCI will allow other applications to link easily to Mosaic in this way

- Server-side example

Here we looked at an example of a server set up specifically to provide access to scientific data: NOAA’s Climate Diagnostics Center (**http://www.cdc.noaa.gov/**). Some noteworthy links that we looked at:

Software support and affiliations (provides access to information about software support, and even a means to get the software)

Map room (consisting of a mix of short-term operational weather products available form the National Meteorological Center)

We also planned, but did not have time, to demo other servers that provide server-side computational support. These servers can perform computational analysis on data, then send the results to the user.

Other interesting examples of scientific data servers:

Pacific Marine Environmental Laboratory
(**http://www.pmel.noaa.gov/pmelhome.html**)

ERIN Database Gateway (**http://kaos.erin.gov.au/database/db.html**)

GeoWeb (**http://wings.buffalo.edu/geoweb/**)

6.7 Challenges and Opportunities for Scientific Information Managers: The NSF/BIR Perspective

Susan Stafford (Division Director, Biological Instrumentation and Resources, National Science Foundation)

The beginning of the 21st century is to biology what the 20th century was to physics. It has been said that the growth of the global Internet and increased computing power have catalyzed entirely novel research methodologies (Brauman, Voss, & Appenzeller (editorial), *Science*, 12 August 1994.) Consider the impacts of the Internet, Gopher, World Wide Web (WWW), and Mosaic upon the manner in which scientific research is conducted and information is disseminated. We are in the midst of an information explosion, coupled with a technological revolution, within the biological and ecological sciences.

This information explosion spans the scale of biological organization from the molecule to the ecosystem, particularly in the fields of molecular biology (sequences and structures), ecology (long-term datasets, spatial data using GIS, remote sensing, etc.), neurosciences (imaging and mapping), and systematics and taxonomy (museums and collections).

The technological revolution has occurred in the areas of data storage (storage media options and capabilities are expanding), visualization (software for molecular- to ecosystem-level modelling, GIS, imagery), high-speed networks (networks are co-evolving with increased speed of the NSFNET backbone), and computation (unexpectedly powerful workstations from the “desktop to the teraflop”).

The laboratory of the 21st century has been called “the anyplace-workplace” --- a ubiquitous computing and communication environment invoking the concept of the collaboratory, which can be considered as “scientists working together *apart*” (i.e., in spite of geographic separation). Researchers in the field will be linked to their offices and labs via wireless communications and will have the capability of accessing, on a real-time basis and from any of these locations, processed data and images. The collaboratory is a richly networked set of resources: people, instruments, and information (databases). The laboratory of the future will be the portal to this resource-rich, digital environment with full accessibility and connectivity among networked resources and people.

The Division of Biological Instrumentation and Resources (BIR) is the *best kept secret* in the Biological Sciences Directorate and will play a major role in the development of the infrastructure for this laboratory of the 21st century. Two programs deserve particular mention: the Database Activities Program and the Research Training Groups. The Database Activities Program has tandem goals of (1) enabling science through database tool/software development and (2) sustaining long-term research infrastructure through database collections and community research resources. Dr. John Porter (jporter@nsf.gov) is the program officer for this program.

The BIO Research Training Groups (RTGs) are new or significantly enhanced multidisciplinary training programs at Ph.D.-granting institutions. They are intended for well-defined research involving appropriate faculty members. They encompass BIO research, but may include other areas as well. Awards include support for undergraduate, graduate, and postdoctoral students, special courses and meetings, student travel, and cost of student research. They are 5-year awards of an average $300K/year, with as much as an additional $200K for equipment. BIR has 29 planned or existing awards in this program. Dr. Gerald Selzer (gselzer@nsf.gov) is the cognizant program officer.

In summary, BIR has several unique, overarching themes. We look for solutions that *scale*, solutions that *transcend domains*, and we take risks! We invite all interested “web-surfers” to visit our programs and find out more about the Division of Biological Instrumentation and Resources on our soon-to-be-released Mosaic home page!

7.0 Appendices

7.1 Site Flashes

Note: With the exception of minor editorial revision in structure, grammar, etc., the following have not been edited for content since they were provided in late September 1994.

7.1.1 AND (Don Henshaw, Mark Klopsch)

This has been an exciting year for the Andrews LTER site. The new on-site laboratory (5000 sq. ft.) was completed last fall and a dedication ceremony was held last November in conjunction with the site’s 40-year anniversary. Plans for an on-site conference center to be built next year are well underway. Sixty tours have passed through the Andrews this past year including groups from grade schools to Oregon Senator Mark Hatfield, as well as numerous researchers and land-use managers. This summer’s REU student program was also very successful.

From an information management point of view, the most exciting news is Fred Swanson’s and other P.I.’s new enthusiasm for the development of our World-Wide Web information server. This has helped our Quantitative Sciences Group in focusing our efforts toward the design and implementation of on-line Internet access. We are hoping to have continued P.I. involvement in this design process. There is strong interest in sharing information and data with the public about Adaptive Management Areas (part of the President’s Northwest Forest Plan which includes the Andrews).

Spatial studies, landscape to regional in scale, continue to grow dramatically. Spatial patterns of land use and their effects on carbon storage, biodiversity, and hydrology are the major areas of emphasis. In connection with these studies, we have developed an atlas of nearly 40 Andrews’ GIS coverages, and can display them using Mosaic. We have also recently been awarded an NSF grant for spatio-temporal analysis at the landscape to regional scales that will support student training over the next five years.

The Andrews is in the process of revamping its climatic and hydrologic measurement program. Two new meteorological stations have been constructed this summer, and we will completely instrument the high elevation station this fall. The other station will be completed next summer. Measurement collection at all stations has been standardized in terms of variables collected and methods used.

All stations are equipped with Campbell data recorders, and all will be networked using radio telemetry within the next year.

New phone lines have been run to the new Andrews headquarters, and we are in the process of setting up our network connection between the Andrews and Oregon State University. The initial installation will use a pair of Netblasers to provide dial-up support for TCP, IPX, and Appletalk communications at 14.4 baud. We will have a dedicated 56-Kb connection with dial-up backup by early next year.

We have also missed Susan Stafford this past year, but expect her return in January. In Susan's absence, our network has grown dramatically, with the number of Unix boxes doubling, and the number of users and PCs increasing by one-third.

7.1.2 ARC (Jim Laundre)

Dave Jones, the previous Arctic LTER data manager, has gone back to school. The new field laboratories were delivered in August. All of the trailers are now green as per Bureau of Land Management requirements. The field season went well.

7.1.3 BNZ (Darrell Blodgett)

Since taking over for Mark Klingensmith at the start of 1994, I have established a World Wide Web server (www.lter.alaska.edu) to serve information to the Internet. We have established Mosaic as the front-end tool for Internet access at our site. We have Mosaic clients on Unix, MacIntosh, and PC platforms. Mark Klingensmith had established metadata forms for researchers to complete and return for documentation of projects and data files. I have used these forms to automate our on-line system to a high degree.

Current projects include: development of Mosaic forms for entering and editing project description and data file description metadata, completion of network installation so users have Internet access from their desktops, development of Mosaic as a front-end interface to a full -featured DBMS (Ingres currently), development of Mosaic forms for data requests and for researchers to grant access to datasets they maintain, resolution of security/data access issues in consultation with researchers, documentation of the on-line system, modification of both project description and data file description forms to improve automation, persuasion of researchers to document and submit data files, maintenance of current versions of Mosaic and viewers on all three platforms, continuation of Unix and DOS/Windows system administration tasks, maintenance of WAIS index of site information, on-line placement of ASCII datasets as they are documented and development of scripts to allow access through a World Wide Web browser, and development of a Mosaic front end to a full-featured DBMS containing a complete catalog of Bonanza Creek data.

7.1.4 CDR (Clarence Lehman)

El Haddi has moved on to a new position with the National Oceanic and Atmospheric Administration. All of us at Cedar Creek LTER wish him well and miss his outstandingly competent and skillful guidance. I am assuming the responsibilities for data management here, now with help from Charles Bristow, a full-time statistician who has recently joined the team.

This year 2 major Cedar Creek studies achieved international recognition, both in the scientific literature and in the popular press. In studying recovery from the drought of 1988-1989, we obtained the first concrete experimental evidence that biodiversity promotes stability and recovery of ecosystems (Tilman and Downing, Nature 367:363-365). Also, in examining implications of the competition-colonization trade-offs inherent in the Cedar Creek data, we uncovered the startling conclusion that habitat destruction is likely to drive the best competitors extinct first, not just the rare and weak species (Tilman, May, Lehman, and Nowak, Nature 371:65-66). Television crews filmed typical operations at Cedar Creek for broadcast on PBS later this year.

To help support this and other scientific work, we advanced several fronts at our facilities. This year, using custom programs we developed and are continuing to develop for HP100 palmtop computers, all data collected in the field were recorded electronically, with no paper data sheets. Data collection was faster, validation immediate, and transcription eliminated. Data files continue to be available on-line, as they have been here for years. In addition, an on-line menu-driven system to promote systematic review and analysis of the data has been added and is undergoing further development.

We have created a suite of utility programs that maintain on-line laboratory notebooks. The important feature is that laboratory notes are tied to associated files of computer data, source programs, and graphs, even as the directory structures surrounding those files are reorganized. These on-line notebooks are now used during the analysis of existing experiments and during theoretical work leading to future experiments.

To enhance the multiprocessing capabilities of our Sparc 10-54, we have written a task management program and associated utilities to focus multiple CPUs dynamically on a single job. This greatly simplifies large-scale simulations.

To make our data manipulation more independent of specific statistical packages, graphics packages, and database systems, we have prepared a set of data management filters that convert columnar data from our database to matrix format for statistical and graphical input, all maintaining machine-independent ASCII format. Miscellaneous programs to display our data, including a program that animates two-dimensional output of plant population simulations, have also been developed.

We have heard that our LTER grant was approved for an additional 6 years of funding. We also initiated, this summer, two major field studies designed to further test the effects of biodiversity on the productivity and stability of grassland ecosystems.

7.1.5 CPR (Tom Kirchner)

As a result of our site visit last autumn, we have been re-evaluating our plan for data management. In particular, we have been working with the investigators to move data into the system in a more timely fashion, and to make data from recently published studies accessible over the network.

We received funding under the LTER supplement to upgrade our Sun 1000 server with a second processor board and additional disk. These are to be used to support a commercial database system.

We had another outside review panel visit our site in August specifically to help provide guidance on our data management system. These reviews were beneficial in helping focus our attention on areas that need improvement, and in helping establish with our investigators the importance with which data management is viewed within the LTER community.

7.1.6 CWT (Gil Calabria)

During this past year, data management personnel have continued their efforts to further automate the data archival process and to educate P.I.’s and graduate students on the usage of the Coweeta information systems.

Several shell scripts and SAS programs have been written to fully automate the archival of micro-climate datasets for both gradient and gap studies. These programs take Campbell data-logger or multiplexer data as input, decompose them into their respective site/plot locations, and convert them to SAS dataset format. This allows the data manager to perform basic range checks and to view newly collected data graphically with minimal effort. The entire process can be done now within a few hours and the data are promptly returned to the P.I. for further quality assurance and analysis.

Updates to the Coweeta LTER site bibliographic database have continued. Our Gopher server currently offers an up-to-date list of all publications, dissertations, and theses spanning more than 50 years of research. This database is now updated quarterly. In addition, with the help of our new data-entry personnel, the data-entry process for our ongoing research is also on schedule and promptly updated.

Taking the suggestions of our site reviewers, we have initiated training sessions to provide P.I.’s and graduate students with an understanding of all of the services provided by our systems. These sessions included training on Unix, networking, SAS, and ArcView, and they were accepted with enthusiasm by all participants. Hence, with this year's technical supplement two additional workstation will be added to our lab to increase the involvement of our research group.

In addition, following suggestions from Dr. Mark Harmon and the LTER Executive Committee, we have adopted an internal peer review committee to aid in the documentation process. Here P.I.'s from different study groups review new data documentation forms prior to final archival. This scheme promotes inter-study communication and better description of data.

We have recently received our Sun SPARC-1000 server. This is a true server-class machine. It has been configured as both a file server and network server, and all of the services now offered by our SPARC-2 "server" are currently being migrated to this machine. Finally, we are currently developing a hypermedia environment using NCSA Mosaic clients and the HTTPD daemon. This project is an attempt to link raw data and metadata with their respective supporting publications and P.I. information. A small demonstration was presented during our internal annual meeting, and it was embraced with enthusiasm by the Coweeta site. As a result of this demonstration, additional ideas have arisen and most P.I.'s seem eager to take advantage of this environment.

Additions made to the Coweeta GIS include evergreen understory and disturbance history coverages. These coverages describe a series of natural and man-made disturbances occurring at the Coweeta Basin. Other news includes the departure of Kurt Saari, our GIS Manager. Kurt is now working with the South Florida Water Management District. We have welcomed David Gould to fill this position. David has a background in environmental sciences and geography, and we are convinced that he will be a great addition to the project.

7.1.7 HBR (Cindy Veen)

We were finally able to solve logistical problems and get the Hubbard Brook Gopher "up and running" this year. There are a few minor changes to be make to make things less confusing to the user.

A paper entitled "Structure and Function of the Hubbard Brook Data Management System" was published in the March, 1994 issue of the Bulletin of the Ecological Society of America.

Another task this year has been to migrate our data backup system from 9-track tapes to Bernoulli disks.

We also continue to add sample collections to our sample archive system. To date, we have close to 22,000 samples archived. We were fortunate this year, with the help of Rich Boone from Harvard Forest, to add to our archives, 1400 forest floor samples collected in 1962. A number of cooperators have begun to subsample from the archives.

7.1.8 HFR (Richard Lent)

There has been expansion of population ecology research at the Harvard Forest. We are working on better access to the Internet and are expanding the computer infrastructure by increasing the number of workstations and hard-disk storage.

7.1.9 JRN (Barbara Nolen, Kevin La Fleur)

Here at the Jornada LTER site (JRN) we have restructured our data management team. Kevin La Fleur is now the research data manager, in charge of all research data done on the Jornada for the LTER. Barbara Nolen is the spatial data manager, and oversees all the GIS and remote sensing data.

In April of this year we hosted the Fourth Annual Friends of the Jornada Symposium, which proved to be a successful coming together of range scientists and ecologists who have or are planning on conducting research within the Jornada study area as well as other interested parties.

7.1.10 KBS (Lolita Krievs)

During the past year, funds from the LTER supplement and matching university funds, have allowed the KBS LTER to significantly increase its GIS/GPS capabilities. We have upgraded our 6-channel GPS community base station receiver to a 12-channel receiver and have purchased a new GPS field receiver which advertises "sub-meter accuracy". Installation of a real-time differential radio link system, which will allow us to process our GPS data "on the fly", is scheduled for mid October. Development of a "Balloon Based Remote Sensing System" (BBRSS) for assessing vegetation composition and cover across individual field plots is also in the works. Color-IR aerial photographs of KBS properties were shot 24 July 1994 and are currently being reviewed. IDRISI GIS software and Arc/Info's Digital Chart of the World are waiting to be installed.

October promises to be a busy month.

7.1.11 KNZ (John Briggs)

The Konza Prairie LTER data management team spent a large portion of this past year in making the Konza Prairie LTER database available over the Internet. The database has always been available electronically to local investigators over our Novell network, but with more of our investigators at institutions other than Kansas State University, we felt it was time to place more of our data in an environment that is more accessible over the Internet. In the past, some of our climate data, publication records, and soil moisture measurements were available over the Internet using a SQL interface (Oracle). However, our investigators were not comfortable with SQL and maintaining the system was very time consuming. In addition, it was difficult to place our metadata into the Oracle system. Thus, we have started using Gopher and Mosaic as tools to browse and use our database. Based upon records of past data requests, we have prioritized which of our datasets to place on-line. At present, we have 20 different datasets (over 100 data files) on-line. These datasets represents over 95% of all data requests over the past 10 years. In addition to the datasets, we have also made available on-line the metadata. This includes the data documentation (format, units, etc.) and parts of our "Methods Manual". This document describes in detail exactly where (including maps) and how each dataset was collected. This system should get an adequate test as, this next year, our LTER group is in the process of synthesizing the LTER effort on Konza Prairie over the past 13 years.

Other data management news includes the departure of Dr. Haiping Su. Haiping took a job with Argonne National Lab in June 1994 as a remote sensing scientist. He was with the Konza Prairie LTER since 1991. We hired Mr. Larry Ballou in mid-June. Larry's duties include: supervision of the Novell network and maintaining the GIS lab. At the same time, we have signed an agreement with the Computing and Networking Department to maintain our Sun workstations.

7.1.12 LUQ (Eda Melendez)

Our LTER renewal proposal was approved, so we will meet the new century as an LTER site. Three other proposals were also funded, two of which involve NASA and LTER (NASA/IRA, an island-wide land-use project, and NASA/EPSCoR, considering global change, both of which involve other governmental and educational agencies as well). This mean that "tons" (maybe terabytes) of data are being or will be generated.

The Remote Sensing Laboratory personnel have done an enormous amount of work in obtaining DIG (digitalized) files as well as imagery from other sources, generating user-interfaces to access the data, and upgrading hardware. We obtained USGS Soil and DLG (digital line graph) maps, Puerto Rico Department of Natural Resources land-use DIG files, U.S. Forest Service Caribbean National Forest soils, roads, trails, and stand maps, and NASA ER2 overflight island-wide imagery. We still do not have good TM imagery, the last one being 1986. We have, however, ordered SPOT imagery of the whole island under the state-wide SPOT View program. Regarding LTER data, the lab developed a menu-driven user interface to access a tree map for our 16-ha grid plots (the El Verde Big Grid). The lab has a Sun workstation (with approximately 5 gigabytes of storage), which is our "gateway to the world", and all the accessible peripherals (plotters, laserjet, drives, etc.) we needed. The lab was involved in the NASA/IRA proposal and we have just submitted another.

Another achievement from this lab was developing “Spatial Metadata”, copies of which were displayed at this meeting in Seattle. Our remote sensing investigator, John Thomlinson, is very interested in receiving feedback on these forms, as well as collaborating on an effort to develop minimum standards for this kind of documentation.

Our site now can say that is “connected” since we can all access our gateway from our desktops. There will be the time when our LAN will be connected as well. We now have a Gopher server and are Mosaic users. The rainfall data from El Verde are available as ASCII files. We still have to decide which other datasets and LTER site related files (e.g., data management policies) we want to make accessible to the public through these resources.

7.1.13 MCM (Jordan Hastings)

MCM is focused on the integration of a suite of software to facilitate access to its data and information. The intent is to develop a user-seductive interface, with Mosaic and GIS as prime components, which serve the entire project community as an electronic “laboratory without walls” for data and information browsing, analytic/synthetic manipulations, and collaborative work. NSF refers to such facilities as “collaboratories”.

During autumn of 1993 and spring of 1994, MCM evolved a series of progressively more detailed plans for its collaboratory effort, designated SOLA, for Science On-Line, Antarctica. SOLA now includes the University of Nevada, Reno (Departments of Computer Science and Geography) and five industrial partners: Claris, Environmental Systems Research Institute, ProCite, Space Biospheres Ventures (Biosphere 2), and StatSci Division of MathSoft. Based on this “bootstrap” work, in June, MCM received a NSF/EPSCoR Small Meeting grant to bring the parties face-to-face; and on 1-2 August, a successful SOLA-1 workshop was hosted at the Desert Research Institute in Reno. The parties committed to develop a formal collaboratory proposal for NSF consideration by June 1995.

7.1.14 NET (Rudolf Nottrott)

The past year has seen many additions to the LTERnet network support system. We have added new capacity and functionality in high-capacity data storage, connectivity, information services, and inter-site data access.

To improve Internet connectivity for scientists in the field and on the road, we have switched dial-up access from Sprintnet’s X.25-based service to a dial-up terminal server that supports common network protocols (SLIP, PPP, Xremote), in addition to simple modem connections.

The LTERnet Field Connectivity Pack, a package of several public domain and shareware programs preconfigured for use with LTERnet, provides the basic software required to set a dial-up IP connection. The Connectivity Pack is available at LTERnet (Gopher LTERnet.edu, Software for Ecologists, FieldPack) and on floppy disk (by request). The package, is based on Trumpet Winsock as the network protocol stack (SLIP), and includes several useful programs for work on the Internet (FTP, Telnet, Mosaic, Gopher, ping, etc.). Other Internet client programs (notably Eudora for email) also work over SLIP, but have not been included to make the pack fit on a single floppy disk. Researchers from 13 LTER sites have made use of the dial-up access since its implementation in February, and the number of monthly calls has risen to over 2000.

In a related development, of the approximately 780 people now listed in the LTER personnel directory, 90% percent have valid email addresses.

Beginning with the 1993 All Scientists meeting, we have now developed the capability to operate a “mini LAN” consisting of a workstation with one or more attached PCs, that can operate stand-alone or linked to the Internet by SLIP. This setup should be useful for holding workshops or demonstrations in locations not directly connected to the Internet. The “mini LAN” has already been used for this purpose during workshops on global connectivity at the INTECOL Congress and the ILTER Steering Committee meeting at Rothamsted, U.K.

We have implemented a WWW server that contains all of our original Gopher server information, plus more, and also provides an interactive map of the LTER network that provides click-and-point access to other LTER site servers. The server is at URL **http://LTERnet.edu/**. New information and data on the server include the catalog of remotely sensed images (data and metadata), a prototype catalog of datasets with data and complete metadata attached (based on the data of the NIN site archived at LTERnet), an expanded all-site bibliography, general information on LTERnet functions (electronic supplementation of inter-site research), information on emerging international LTER activities (ILTER), and more.

7.1.15 NTL (Barbara Benson)

We are improving access to information for researchers in a number of ways: 1) direct end-user access to Oracle databases, 2) site Gopher and World Wide Web (WWW) servers, and 3) enhanced connectivity of our field station. During the past year, we have acquired Oracle Data Browser, an application which allows researchers to access data in the NTL-LTER Oracle database through a “point-and-click” type interface and to export the data into other applications for analysis. We were also able with the cooperation of the Konza Prairie LTER data manager to use Oracle Data Browser to access the Konza Oracle database. Researchers will now be able to explore Oracle databases in an interactive mode and to make linkages between diverse data tables.

We have created a Gopher server for NTL-LTER on our Sun workstation and have made the server publicly accessible. Currently, the Gopher contains a site description, a personnel directory, the NTL-LTER bibliography, data, and metadata. We chose the NTL-LTER fish data as our first data for Gopher. After discussion, we decided to put some derived fish data on-line rather than the raw data. The derived data aggregate information in useful forms such as length frequencies and also have been modified by correction algorithms such as standardizing to a common unit of sampling effort. We also put the fish sampling protocol on the Gopher server.

A NTL-LTER WWW server is under construction. We liked the map link to information about specific sites that Rudolf Nottrott had set up on the LTERnet WWW server. We have created an image showing our seven study lakes identified by lake name. When the user selects the lake name, he/she is linked to further information about the lake such as morphometric characteristics, summary chemical parameter values, and species lists. The WWW server will also provide a gateway to datasets in the NTL-LTER Gopher. We are designing an information system that provides people with usable information at varying levels of detail.

Installation of the ethernet wiring at our field station, Trout Lake Station, is proceeding in preparation for an Internet connection. We have received funds through the 1994 technical supplement and from the University of Wisconsin-Madison to purchase the necessary hardware and software to establish an ethernet local area network at Trout Lake and a 56-Kbaud dedicated phone line between Trout Lake and the closest Internet node.

The data management committee for NTL-LTER met for a one-day retreat at the field station. Many aspects of data management were reviewed and goals were set for the coming year. Three major goals are (1) to incorporate more core data into the Oracle database, (2) to acquire several existing regional datasets, and (3) to complete the development of the NTL-LTER WWW information server.

7.1.16 NWT (Rick Ingersoll)

We now have 92 datasets archived in a single directory on our workstation and each of these files is identical in structure, in ASCII format, and includes the metadata. After exploratory navigation of the other LTER site Gophers and consultation with site data managers, the NWT LTER Gopher was installed in July of 1994. The data for 54 of the datasets are available through the NWT LTER Gopher; metadata for all 92 of the datasets are accessible through Gopher.

We continue to retroactively apply recently formalized rigorous protocols to pre-centralization datasets, particularly our climate data. We are also trying to automate a portion of the quality control for electronically collected data.

During the autumn of 1993, a subnivean shelter was installed on the Saddle. Meteorological instrumentation was installed there to quantify energy and mass exchanges of an alpine snowpack. This instrumentation consists of 1) a suite of radiometers to measure incident and reflected shortwave radiation, down- and up-welling longwave radiation, net radiation, and to separate the direct and diffuse components of shortwave radiation; 2) temperature, relative humidity, and wind speed sensors at 3 levels to permit calculation of turbulent fluxes of sensible and latent heat using aerodynamic formulae; and 3) a thermocouple ladder to measure energy fluxes within the snowpack and soil using temperature gradient methods. Usable data were collected beginning in the spring of 1994.

A 60-m long, 2.8-m high snowfence was installed on the Saddle in the autumn of 1993. Soil temperature measurements at 2 depths (0 and 15 cm) at this site have been recorded weekly to monthly at this site since the autumn of 1992. Within the next few weeks sophisticated probes will be installed that will enable temperature measurements at 10-cm intervals over the air-snow-soil continuum.

We have purchased a Sun SPARC 20 that will be devoted primarily to modelling. We also have purchased a number of X-terminals. Within the coming months the Mountain Research Station will be ethernetted and we will be able to say good-bye to “sneakernet”!

7.1.17 PAL (Karen Baker)

The third field season from September 1993 to March 1994 was completed with continuation of oceanographic, krill and bird surveys. In addition to the weekly sampling program at Palmer Station, there were two major ship cruises during which core measurements were made off the Antarctic Peninsula.

1)A “lteraug93” austral spring cruise on the established LTER grid was followed by a similar fall cruise, “ltermar93”, to investigate responses of the ecosystem to winter ice conditions. These paired cruises, conducted once within each six-year funding period, aim to determine the effect on the ecosystem of winters of very different annual sea-ice conditions.

2)The “lterjan94” cruise was the second annual cruise of the PAL LTER and resampled stations on several established transects.

Data forms were provided for at sea use on Macintosh-, PC-, and Unix-formatted disks during these two cruises. As a result, much documentation of datasets was accomplished in “real-time”. Consideration of possible procedure manuals is underway as techniques become standardized. Use of two Unix workstations at sea was successful and productive but will continue on a limited basis due to difficulty and expense of maintaining trained personnel in the field.

Preparations are underway for the fourth field season and the “lterjan95” annual cruise. Plans have begun for the installation of a second automatic weather station (AWS) at Hugo Island in addition to the first one on Bonaparte Point. Inquiries continue in an effort to collect historical weather data.

Conversion was made from the on-line documentation browse program to a Gopher site implementation. A Mosaic home page was installed and various formats tested. Mosaic has provided a useful method for sharing graphical information between principal investigators. A new 2.0-gigabyte disk replaced older hardware for data storage. Throughout the last year, plans have been made for interfacing (with the ship and station) computer upgrades and the new ship data acquisition system (DAS).

A LTER Executive Committee Meeting and a Steering Committee Meeting in December 1993 preceded the Palmer LTER site review. The site review was held in two locations: Santa Barbara (5-7 February), where the emphasis was on data management, and Palmer Station in the Antarctic (8-17 February).

7.1.18 SEV (James Brunt)

The data management group at the Sevilleta met in “retreat” last winter to develop some strategies for meeting the goals of the Sevilleta LTERII proposal, regionalization proposal, and improving communications among the P.I.’s and staff. Since then we have been actively pursuing many of these directives. We have implemented both Gopher and WWW servers and are continually increasing the amount of information available there. We are also completing the archival process for our first six years worth of data. Our supplement proposal this year focused on networking and computerizing the laboratory analytical equipment. We were successful and have begun that process. The Sevilleta Research Station was completed mid-summer and now has the capacity for 48 researchers, the phase III development included the construction of a computer library building with 23 visiting scientist carrels. Each equipped with a Macintosh, PC, or X-terminal. In addition, the new residences are all networked with 10BaseT and terminal server access. The sun photometer that was installed this spring as part of the NASA collaboration continues to function without too much input from us, although we would like to find a way to download the data locally, in addition to the satellite up-link. We’ve also taken delivery of a number of remotely sensed images as part of that collaboration and have completed geometric and radiometric corrections on them.

7.1.19 VCR (John Porter)

The past year has seen a continued focus on the development of VCR network capabilities. The principal change has been the creation and enhancement of a World Wide Web (WWW) hypertext information server. Currently this server plays two primary roles. First, it serves as a gateway to the mass of information stored in our Gopher information server over the past several years. Secondly, it provides a way of presenting longer documents which integrate text and graphics, such as our recent renewal proposal. These documents are broken into chunks and linked with hypertext links for rapid access. Figures are scanned or converted directly to GIF files for inclusion with the document.

We have also used our server to meet larger needs. We served as the coordinating site for the creation of the LTER Regionalization document. This involved text-to-hypertext conversions of emailed text and graphics format conversions. We have also been serving as a home site for information on Hungarian LTER efforts until they can get on the Internet themselves.

John Porter, along with Rudolf Nottrott (NET) and James Brunt (SEV) participated in three demonstrations of use of the Internet to facilitate ecological research. These took place at the National Science Foundation (for an audience from a variety of federal and non-profit organizations), at the International Congress of Ecology, and at the International Long-Term Ecological Research Steering Committee Meeting.

This has been a very big year for proposal writing at the VCR LTER! Writing our renewal proposal required communicating text on an almost daily basis between scientists at 7 institutions in 4 states. It was finally wrapped up with a marathon group writing session involving approximately 14 scientists and 5 PC or Macintosh computers, each with a different section of the proposal on it. By using FrameMaker software on all of these systems, it was easy to exchange files among those 2 platforms and Unix workstations running FrameMaker. The completed proposal (including graphics) was then converted to hypertext for display on our WWW server. This process was repeated on a smaller scale for our augmentation proposal and for several non-NSF proposals to supplement different parts of the VCR research program.

We have also been active in developing new data systems. Dave Krovetz successfully adapted Onset Computer’s HOBO data logger to new sensors for use as a simple, low cost, electronic water-level monitoring system. Interest in this system was so high that Onset built on Dave’s work to come out with a whole new line of water pressure sensors. We are still working on developing all the supporting system (metadata, QA/QC) for a collection of some 21 loggers operated by several different investigators.

A major effort this summer was creation of high-accuracy topographic data layers of critical study sites using laser total station and GPS technologies. The resulting information has been added to our GIS system for use by all researchers working at the site.

We are also about to become the primary data provider for the Northampton County, VA Geographical Information System under a contract funded by NOAA. This project will involve the creation of numerous data layers in Arc/Info to support creation of a special area management plan. In addition to using Arc/Info, we are planning on using Mosaic as hypertext metadata access tool.

John Porter has been active in participating in a variety of workshops concerning metadata. These included the NASA System Interoperability Workshop (also attended by Rudolf Nottrott), the IEEE Metadata Workshop, and a workshop focusing on interoperability between herbarium databases. His conclusion is that nobody has found “the answer” but lots of people are looking! In fact, the efforts so far by LTER data managers are better developed than most other groups.

There have been a few personnel changes. John Porter is now working only part-time with the VCR LTER. He is now splitting his time between the VCR and the National Science Foundation, where he is overseeing the Database Activities Program. David Richardson, who began with the VCR doing vegetation mapping, has taken over most of the day-to-day operations. Dave Krovetz continues to oversee the electronic monitoring program.

7.2 Agenda, LTER Data Managers Business Meeting

Wednesday, 21 September 1994, University Inn Board Room

7:00-10:00 PM - Evening Reception

Mixer

Site “Bytes”

Finalize Agenda

Thursday, 22 September, Univ. of Washington, Husky Union Building, Room 106B

8:00 AM - Old Business (Facilitator: James Brunt)

Recommended Technological Capabilities

Committee Funding and Chair Selection

Open to the Floor (from email reports)

Work Session for Old Business (if needed)

10:45 AM - Speakers and Discussion

Mark Harmon (AND) - Data Managers’ Facilitation of Inter-site Research

11:45 AM - Working Lunch

1:00 PM - Speakers and Discussion (continued)

Susan Stafford - Perspectives from NSF

2:00 PM - New Business (Facilitator: John Briggs)

Response to CC Instructions on Metadata Standards and On-Line Datasets

- Summary of the Current Status
- Considerations that must be Addressed in Dealing with these Charges

3:00 PM - Work (Facilitator: Rick Ingersoll)

Working Group Reports

Next Data Managers Meeting

- format
- organizing committee (dmmoc) status

Wrap-Up and Meeting Report (completion before October CC meeting)

Last-Minute Coordination of the “Open” Meeting

7.3 List of LTER Data Managers Business Meeting Participants

Karen Baker (PAL)

UCSD/SIO, A-018, La Jolla CA 92093-0218

LTERnet address: KBaker@LTERnet.edu, E-Mail: karen@crseo.ucsb.edu

Phone: (619) 534-2350, FAX: (619) 534-2997

Barbara J. Benson (NTL)

University of Wisconsin-Madison, Center for Limnology, 680 N. Park Street, Madison WI 53706

LTERnet address: BBenson@LTERnet.edu, E-Mail: bbenson@macc.wisc.edu

Phone: (608) 262-2573, (608) 262-3088, FAX: (608) 265-2340

Caroline S. Bledsoe (NET)

University of California-Davis, Department of Land, Air and Water Resources / Hoagland Hall, Davis CA 95616

LTERnet address: CBledsoe@LTERnet.edu, E-Mail: csBledsoe@ucdavis.edu

Phone: (916) 752-0388, FAX: (916) 752-1552

Darrell Blodgett (BNZ)

University of Alaska, Forest Soils Laboratory, 305 O’Neill Building, Fairbanks AK 99775-0740

LTERnet address: DBlodgett@LTERnet.edu, E-Mail: blodgett@taiga.lter.alaska.edu

Phone: (907) 474-7036, FAX: (907) 474-7439

John M. Briggs (KNZ)

Kansas State University, Division of Biology, Ackert Hall, Manhattan KS 66506-4901

LTERnet address: JBriggs@LTERnet.edu, E-Mail: jmb@andro.konza.ksu.edu

Phone: (913) 532-6629, FAX: (913) 532-6653

James W. Brunt (SEV)

University of New Mexico, Department of Biology, Castetter Hall, Albuquerque NM 87131-1091

LTERnet address: JBrunt@LTERnet.edu, E-Mail: jBrunt@sevilleta.unm.edu

Phone: (505) 277-9342, (505) 277-9372, FAX: (505) 277-0304

Gil N. Calabria (CWT)

University of Georgia, Institute of Ecology, Athens GA 30602

LTERnet address: GCalabria@LTERnet.edu (or GCalabri), E-Mail: gil@sparc.ecology.uga.edu, gcalabri@uga (Bitnet)

Phone: (706) 542-5691, (706) 542-2968, FAX: (706) 542-6040

Harvey Chinn (NET)

University of California, Division of Environmental Studies, Davis, CA 95616

LTERnet address: harvey@LTERnet.edu, E-Mail: harvey@ice.ucdavis.edu

Phone: (916) 752-6300, FAX: (916) 752-3350

David Gould (CWT)

University of Georgia, Institute of Ecology, Athens GA 30602-2202

LTERnet address: DGould@LTERnet.edu, E-Mail: david@sparc.ecology.uga.edu

Mark E. Harmon (AND)

Oregon State University, Department of Forest Science, Forestry Sciences Lab 020, Corvallis, OR 97331-7501

LTERnet address: MHarmon@LTERnet.edu, E-Mail: harmon@fsl.orst.edu

Phone: (503) 750-7333, FAX: (503) 737-1393

Jordan T. Hastings (MCM)

Biological Sciences Center, Desert Research Institute, P.O. Box 60220, Reno, NV 89506

LTERnet address: JHastings@LTERnet.edu, E-Mail: jordan@maxey.unr.edu

Phone: (702) 673-7445, FAX: (702) 673-7485

Donald L. Henshaw (AND)

USDA Forest Service, Pacific NW Station, 3200 SW Jefferson Way, Corvallis OR 97331

LTERnet address: DHenshaw@LTERnet.edu, E-Mail: henshaw@fsl.orst.edu

Phone: (503) 750-7335, FAX: (503) 750-7329

Rick Ingersoll (NWT)

University of Colorado, INSTAAR, Campus Box 450, Boulder CO 80309-0450

LTERnet address: RIngersoll@LTERnet.edu (or RIngerso), E-Mail: ricki@culter.colorado.edu

Phone: (303) 492-4771, (303) 492-0566, FAX: (303) 492-6388

Thomas B. Kirchner (CPR)

Colorado State University, Natural Resource Ecology Laboratory and Department of Range Science, Fort Collins CO 80523

LTERnet address: TKirchner@LTERnet.edu (or TKirchne), E-Mail: tom@chloris.nrel.colostate.edu

Phone: (303) 491-1986, FAX: (303) 491-1965

Mark Klopsch (AND)

Oregon State University, Department of Forest Science, Forestry Sciences Lab 020, Corvallis, OR 97331-7501

LTERnet address: MKlopsch@LTERnet.edu, E-Mail: klopsch@fsl.orst.edu

Phone: (503) 750-7339, FAX: (503) 737-1393

Lolita Krievs (KBS)

Michigan State University, W.K. Kellogg Biological Station, Hickory Corners MI 49060-9516

LTERnet address: LKrievs@LTERnet.edu, E-Mail: krievs@kbs.msu.edu

Phone: (616) 671-2214, FAX: (616) 671-2351

James Laundre (ARC)

Marine Biological Laboratory, Ecosystems Center, Woods Hole MA 02543

LTERnet address: JLaundre@LTERnet.edu, E-Mail: jimL@lupine.mbl.edu

Phone: (508) 548-3705 x476, FAX: (508) 457-1548

Kevin La Fleur (JRN)

New Mexico State University, Department of Biology, Box 30001, Dept. 3AF, Las Cruces NM 88003

LTERnet address: KLaFleur@LTERnet.edu, E-Mail: klafleur@nmsu.edu

Phone: (505) 646-7918

Clarence Lehman (CDR)

c/o University of Minnesota, Department of Ecology, Evolution, & Behavior, 100 Ecology Building, 1987 Upper Buford Circle, St. Paul, MN 55108-6097

LTERnet address: CLehman@LTERnet.edu, E-Mail: lehman@lter.umn.edu

Richard A. Lent (HFR)

Harvard University, Harvard Forest, PO Box 68, Petersham MA 01366

LTERnet address: RLent@LTERnet.edu, E-Mail: harvard.forest (Omnet)

Phone: (508) 724-3302, FAX: (508) 724-3595

Eda C. Melendez (LUQ)

Terrestrial Ecology Division, P.O. Box 363682, San Juan PR 00936

LTERnet address: EMelendez@LTERnet.edu (or EMelende), E-Mail: e_melendez@upr1.upr.clu.edu

Phone: (809) 767-0350, FAX: (809) 758-0815

Barbara Nolen (JRN)

New Mexico State University, Department of Biology, Box 30001, Dept. 3AF, Las Cruces NM 88003

LTERnet address: BNolen@LTERnet.edu, E-Mail: bNolen@nmsu.edu

Phone: (505) 646-4465, FAX: (505) 646-5665

Rudolf W. Nottrott (NET)

University of Washington, College of Forest Resources, AR-10, Seattle WA 98195

LTERnet address: RNottrott@LTERnet.edu (or RNott), E-Mail: rNott@lternet.edu

Phone: (206) 543-8492, FAX: (206) 685-0790

John H. Porter (VCR)

University of Virginia, Department of Environmental Science, Clark Hall, Charlottesville VA 22903

LTERnet address: JHPorter@LTERnet.edu, E-Mail: jhp7e@virginia.edu

Phone: (804) 924-7761, FAX: (804) 982-2137

Gregory A. Shore (SEV)

University of New Mexico, Department of Biology, Biology Annex, Albuquerque NM 87131

LTERnet address: GShore@LTERnet.edu, E-Mail: gShore@sevilleta.unm.edu

Phone: (505) 277-2109, FAX: (505) 277-5355

Susan Stafford (NSF)

Division Director, Biological Instrumentation and Resources, National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230

LTERnet address: SStafford@LTERnet.edu, E-Mail: sstaffor@nsf.gov

Phone: (202) 306-1470, FAX: (703) 306-0356

Cindy Veen (HBR)

USDA Forest Service, Forestry Sciences Laboratory, P.O. Box 640, Durham NH 03824

LTERnet address: CVeen@LTERnet.edu, E-Mail: c_Veen@unhh.unh.edu, or S24L06A (USFS)

Phone: (603) 868-5692, FAX: (603) 868-1538

7.4 Agenda, Open Workshop: The Management of Spatial Data and Inter-site Data Access in the Ecological Sciences

Friday, 23 September 1994: Management of Spatial Data

8:00-8:15 AM - Introduction (welcome, overview of agenda and goals)

8:15-8:30 AM - Overview of LTER and LTER Data Management (James Brunt)

8:30-9:30 AM - Introductions and Overview Presentations (non-LTER) by Participating Groups

9:30-10:15 AM - Jim Frew:

“Data Management Storage: Today and Tomorrow”

10:15-10:45 AM - Break

10:45-11:30 AM - Kenn Gardels:

“GeoData Management Using Mapping, Imaging, and DBMS Software Tools”

11:30-12:00 AM - Nancy Tosta:

“Federal Geographic Data Committee Standards for Spatial Metadata”

12:00-1:00 PM - Lunch

1:00-1:45 PM - Bill Michener:

“Spatial Metadata for the Environmental Sciences”

1:45-2:00 PM - Working Group Organization

2:00-3:30 PM - Working Group Meetings

Network of Networks

Standards Development

Exchange of Spatial Data

User Access and Image Catalogs

Proprietary Issues

3:30-4:00 PM - Break; Report Organization; Demonstrations

4:00-5:00 PM - Working Group Reports

Evening - Group Dinner

Saturday, 24 September1994: Inter-site Data Access

8:00-8:15 AM - Organization; Overview; Announcements

8:15-9:00 AM - David Fulker:

“Models, Data Formats, and Software Tools Applicable to Inter-site Data Exchange”

9:00-9:45 AM - Mike Folk:

“Integrating Modern User Interfaces and Scientific Data Formats”

9:45-10:15 AM - Susan Stafford:

“Challenges and Opportunities for Scientific Information Managers: The NSF/BIR Perspective”

10:15-10:45 AM - Break; Demonstrations

10:45-11:00 AM - Jim Quinn:

“Non-Spatial Data Issues”

11:00-12:00 AM - Overview of LTER Network Office

12:00-1:00 PM - Lunch

1:00-1:30 PM - XROOTS Project Overview

1:30-3:30 PM - Working Group Meetings

Sources of Funding for Database Development; Research Areas in Scientific Data Management

Metadata Standards and Exchange

Implementing On-Line Information Systems Using Generic User Interfaces

Interfacing Generic Network Tools with Other Software

3:30-4:00 PM - Break; Report Organization; Demonstrations

4:00-5:00 PM - Working Group Reports; Wrap-Up

7.5 List of Participants in 23-24 September Sessions

The following people (in addition to the participants listed in 5.3) attended the 23-24 September sessions:

Scott Chapal

Database and Network Manager, J.W. Jones Ecological Research Center

John Faundeen

Senior Information Scientist, Hughes STX (c/o U.S. Geological Survey)

Mike Folk

Head, HDF Group, National Center for Supercomputing Applications, University of Illinois (Urbana/Champaign)

E-Mail: mfolk@ncsa.uiuc.edu

James Frew

Specialist, Institute for Computational Earth System Science, University of California (Santa Barbara)

E-Mail: frew@icess.ucsb.edu

David Fulker

Director, Unidata Program Center, University Corporation for Atmospheric Research

Kenn Gardels

GIS Research Specialist, CEDR - 390 Wurster Hall #1839, University of California, Berkeley, CA 94720-1839

E-Mail: gardels@ced.berkeley.edu

Phone: (510) 642-9205, FAX: (510) 643-5571

Merilyn J. Gentry

Project Scientist, Oak Ridge National Laboratory Distributed Archive Center

Bruce Gritton

Monterey Bay Aquarium Research Institute

John J. Helly

San Diego Supercomputer Center

Marge Holland

Chief, Integration and Assessment Team, Environmental Monitoring and Assessment Program, Research Triangle Park, NC 27711

E-Mail: HOLLAND.MARGE@epamail.epa.gov

Jeff Jefferson

Information Manager, Baruch Institute, University of South Carolina

George Lienkaemper

GIS Coordinator, Pacific NW Station, USDA Forest Service, 3200 SW Jefferson, Corvallis, OR 97331

E-Mail: geo@fsl.orst.edu

Robert MacArthur

Network Coordinator, College of Agriculture, University of Arizona

Arthur McKee

Director, H.J. Andrews Experimental Forest

David M. Mark

Professor of Geography, State University of New York (Buffalo), National Center for Geographic Information and Analysis

Linda May

Land Margins Ecosystem Research

William K. Michener

Joseph W. Jones Ecological Research Center, Route 2, Box 2324, Newton, GA 31770

E-Mail: wmichene@longleaf.jonesctr.org

Phone: (912) 734-4706, FAX: (912) 734-4707

Michelle Murillo

Oregon State University

Jim Quinn

Division of Environmental Studies, University of California (Davis)

Bob Shepanek

Information Management Coordinator, Environmental Monitoring and Assessment Program

Nancy Tosta

Staff Director, Federal Geographic Data Committee, 590 National Center, Reston, VA 22092

E-Mail: ntosta@usgs.gov

Phone: (703) 648-5725, FAX: (703) 648-5755

Robb Turner

Project Scientist, Oak Ridge National Laboratory Distributed Archive Center

Leonard J. Walstad

Professor, Chairman of GLOBEC Data Management Committee, Horn Point Environmental Laboratory, University of Maryland