

# PROCEEDINGS OF THE 1990 LTER DATA

## MANAGEMENT WORKSHOP

Editors:

we Michener, S. Stafford, R. Nottrott

Contributors:

D. Henshaw, G. Spycher, S. Stafford, B. Molier, P.  
Adams, A.El-Haddi,

T. Kirchner, G. Calabria, C. V~n, E. Boose, E. H.

Muidavin, Je Gorentz, J. M. Briggs, E. C.  
Meiendez-Coiom, J. Vande Castle, R. Nottrott, we  
Michener, B. Benson, J. Halfpenny, R. Ingersoll, J.  
W. Brunt,

J. Porter

PROCEEDINGS OF THE 1990 LTER DATA MANAGEMENT WORKSHOP

(Held in Snowbird, Utah July 26-28, 1990)

CONTENTS:

I. Executive Summary

II. Introduction

III. General Discussion Topics

A. Review of last year's meeting

B. New directions

C. CEES

D. Data catalog

E. MSI

F. DBMS

G. Standardized data files for intersite exchange of data

H. Archival hardware (optical disks)

I. DOE Research Parks

J. Agenda for 1990 All Scientists Meeting

K. Agenda for 1991 Data Management Workshop

L. Criteria for Data Management Site Evaluation

M. Quality assurance

N. DataBits

0. tidBITS

#### **Iv.** Working Group Reports

A. Climatological databases: two approaches

B. GIS/Remote **Sensin~ata** Management Workshop

C. Outreach activities

D. Proprietary issues

E. Meta-data

F. Networking

Appendix L Attendees

Appendix **II.** Site Reports

### PROCEEDINGS OF THE 1990 LTER DATA MANAGEMENT WORKSHOP

(Held in Snowbird, Utah July 25-27, 1990)

#### I. EXECUTIVE SUMMARY

The LTER Program is unique in that it represents a network. An explicit goal of the LTER program is to study ecological processes that require measurement over long periods of time (years to decades to centuries). The LThR Data Managers recognized these two aspects of the LTER program and devoted their 1990 LTER Data Managers Workshop to the discussion and resolution of several associated problems and challenges in meeting these objectives. This meeting resulted in a series of specific accomplishments and action items:

(1) Decision to develop a proposal for a Symposium focussing on "**Data** Management

Issues for Global Environmental Change for the 1990's: New Tools, New Issues, New

**Challenges."**

(2) Planning for creation of a distributed climate database.

(3) Discussion of mechanisms for establishing a series of data management workshops for the Chinese LTER Program.

(4) Initiated planning for a workshop on data management linkage to GIS/Remote

Sensing activities with an emphasis on documentation and data management standards.

(5) Creation of a short position paper on long-term archival devices (optical disks).

(6) Initiation of an out:reach endeavor to contact other agencies and programs dealing with research information management on a global and regional scale.

(7) Development of a one page summary for investigators relating to minimum

documentation standards for data abstracts.

(8) Criteria for LThR data management site evaluation were examined.

(9) Establishment of an electronic connection between LTERNET and Parkhet (DOE research parks) was proposed and has been implemented.

(10) New topics for 1990-91 Data Bits were identified.

(11) Agreement to search for new funds to connect LTERNET with other groups within the ecological community.

## II. INTRODUCTION

LTER Data Managers (Appendix I) met in **Snowbird**, Utah (July 26-28, **1990**) for their annual workshop, prior to the ESA meetings. The meeting followed the format of the successful 1989 LTER Data Managers Workshop. Initial discussion focussed on specific objectives for the meeting and general topics of interest. Each data manager discussed progress at their site (Appendix **II**). This was followed by a series of smaller working groups sessions where specific topics were examined in greater detail. **~aders**, identified by each working group, were responsible for oral summaries to the group as a whole as well as providing written working group summaries to be included in this report. Drs. John Pfaltz and Walt Conley were invited to participate on the second full day of the meeting. Dr. Pfaltz, University of Virginia, gave a presentation on the differences between Scientific and Commercial Databases, and Dr. W. Conley, New Mexico State University, discussed his experience with the DOE Research Parks in Washington, DC.

This report summarizes the topics discussed in Snowbird, Utah and lays out the 1990-91 action items and agenda for the LTER Data Managers.

## III. GENERAL DISCUSSION TOPICS

### A. REVIEW of 1989 DATA MANAGEMENT MEETING

The 1989 LTER Data Managers meeting was reviewed to provide an overview for people who had not attended the 1989 Toronto Meeting as well as to give a sense of continuity to the activities of the Data Managers. For a detailed report, the interested reader is referred to the formal report presented at the 1989 LTER/Coordinating Committee **~TER/CC**) meeting at Harvard Forest.

Key items included: proposals to create a core data set catalog and a bibliographic database; the establishment of a Task Force (Stafford (AND), Michener (NIN), Porter (VCR), and Brunt (SEV)) to develop future goals, provide liaison between NSF, P1's, **LThRICC**, and the Data Managers; establishment of a quarterly newsletter called DataBits (Porter (VCR)), and a compilation of the historical LThR data management literature (Nottrott (NET)). Recommendation for action and furthe? discussion included recognizing Data Management as a key element of LTER, developing GIS documentation standards with the GIS group, and a Unix system administration workshop. The progress on many of these topics has been very good. The current status of the Core Data set Catalog is "in press", **Da~its** has been published on a regular basis, the Task Force has been the initial point of contact with NSF and the

rest of the LTER community, and a historical file of data management documents has been collated by the Network Office and will be disseminated electronically in the near future. The bibliographic proposal did not receive the same level of support as did the core data set catalogue proposal. Data Management as a sixth core area generated much discussion at the Harvard Forest LTER/CC meeting but did not garner adequate support among the

### 3

majority of PIs although several of the site PIs were in favor of the designation Developing GIS standards with the GIS committee is on-going. And finally the Unix administration workshop proposal was still pending.

## B. NEW DIRECTIONS

A brainstorming session came up with the following list of projects/topical areas for the Data Managers to lead with into the **1990's**. Several ideas were discussed and could be organized around four major areas: (1) Data Management Symposium; (2) **GIS/Remote Sensing** Management workshop; (3) Distributed database for climate and decomposition; and (4) Outreach to international and national communities.

### Data Management Symposium

There was agreement that a follow-up symposium to the Baruch Symposium would be appropriate and timely. Funding will be sought from Dr. R. **Robbins**'s new initiative which supports workshops, colloquia, and symposia. This may evolve into a series of regular symposia on data management topics. William Michener (NIN), James Brunt (SEV), and Susan Stafford (AND) will write the pre-proposal for fall submission.

### GIS/Remote **Sensing** Management Workshop

A GIS/Remote **Sensing** Management workshop or symposium would be useful to develop, as a product, some standards and protocols. This could be best done by linking to national expertise in the GIS/Remote Sensing arenas. John Vande Castle (NET) will pursue this.

### Distributed Network Databases

The concept of distributed versus centralized databases was discussed. Two examples were suggested; a network climate and decomposition databases. The initial project would be for demonstration and prototype development.

A network climate database would include monthly minimum and maximum temperature, and precipitation. Each participating site would set up a separate account that Tom Kirchner (CPR) (or any other site, eventually) would have access to via the network. Tom would write interactive software for retrieval of data from other sites analysis and display. This capability would be limited to SUN Workstations on the Internet. It was suggested that SUN might be interested in providing support for this project.

Another possibility would be to use e-mail access along the lines of a herbarium link demonstrated by John Gorenz (KBS) at the April data management workshop at Kellogg. This would require those sites that provide data to have SQL. It may be that email is a viable common denominator for exchange.

## Outreach Activities

Several outreach activities were discussed. At the network level, this included a televised workshop, a newsletter, short position papers, and demonstrations at individual **~ites** and LTER/CC meetings. Specifically for PIs, it would also be helpful to identify intersite projects and develop ways to improve the access by researchers to the data.

At the national level, a compilation of site methods manuals would be helpful. Rudolf Nottrott (NET) has started keeping a historical evolution file of data management at the LThR sites. These include reports, proceedings, and other writings. Ideally, this would be available in both printed and digital formats. The collection of data management manuals from each site should be included as the next step. Another activity at this level, would be the establishment of a mail forwarding system for the ecological community including biological field stations.

At an international level, we discussed a series of overview workshops on Data Management for international LThR sites. LTER could take the lead in this area.

## C. COMMITTEE on EARTH and ENVIRONMENTAL SCIENCES

The Committee on Earth and Environmental Sciences (CEES) is a multi-agency group which is organized to coordinate research and funding in the U.S. for global climate change. CEES has produced a series of three small format booklets which describe these coordination efforts ("Our Changing Planet" for Fiscal Years 1989, 1990, and 1991) and two large format books, which detail the proposed research by each agency.

The proposed research is organized by seven science elements and is prioritized -a key point. Further, each agency develops its proposed research and funding requests in concert with other agencies - a new and important step for federal agencies, which often work alone.

Much of the new funding is targeted for NASA's space platform, EQS, Earth Observing System, but other areas of science, including ecology are also proposed for increases in funding. This budget is currently being discussed on **Capitol** Hill, but it is uncertain in the current budget climate (national debts, military expenditures in the Middle East, savings and loan fiasco, etc..) what Congress will authorize.

In addition to the CEES plan, a number of U.S. agencies are developing their own plans (which are part of CEES), and these agencies are interested in interacting with the LThR network and with specific sites. Some of these agencies will present the Global Change research plans at the All **Scientists'** meeting in Colorado. For example, the U.S. Forest Service will describe their Forest Health Monitoring program, and EPA will outline EMAP **~nvironmental** Monitoring and Assessment Program). The Department of Energy's Parknet program is already interacting with LThR since many of these sites

## 5

(4) attended the LTER-sponsored GLOBAL Change conference in Denver last November, the result of which was the **'1990's** Global Change Action Plan".

The USGS is also interested in developing a program in hydrology (WEBB, **Water**, Energy and Biogeochemical Balances) with some of the LThR sites.

## D. DATA CATALOG

The **LThR** Data Managers endorsed expansion of the Data Catalog to include site descriptions and maps. The Catalog has grown to over 300 pages because of these additions, plus the fact that sites have more ongoing LThR studies than originally estimated. The Catalog is in press and will be available for fall 1990 distribution. Questions concerning the Data Catalog can be addressed to William Michener & Anne Miller (NIN) and Rudolf Nottrott (NET).

#### On-line version of the LTER Catalog of Core Data Sets

The Catalog entries are now on-line at the LTERNET computer system as ASCII text and in WordPerfect format. The files are located in directory **-ftplcatalog** and are presently stored on a per site basis. More files to be added shortly include a subject index, an investigator index, site abstracts and a reference guide.

Possible access methods include anonymous file transfer using **FTP** - File Transfer Protocol - over the Internet (host lternet.washington.edu, Internet address 128.95.36.1), or dial-in over the phone system (phone **20&543-2115**) using ASCII text transfer or the Kermit program. Automatic mail reply for subsets of the files will be enabled later.

For more information you can obtain the most recent reference guide to the online catalog (including details of access methods), by sending any message to Internet address

**HelpCtlg@lternet.washington.edu** (or Bitnet address **HelpCtlg@LTERNET**).

For more general help on the LTERNET information system, send any message to

**Hel~** lternet.washington.edu (or **Hel~LThRNET** on Bitnet).

In addition to the files that are presently on-line, the Network Office plans to maintain the Catalog, or parts of it (indices, etc.), in a relational database management system based on the SQL query language standard. This will allow easy file exchange and query. Information on the status of this implementation, and other relevant details, are also summarized in the reference guide returned by the '**HelpCtlg**' function.

#### E. MINIMUM STANDARD INSTALLATION (MSI) at LTER SITES

Emery Boose (HFR) collected information on GIS and remote sensing systems at each site as part of a • **'post-MSI'** survey. The results will be discussed at the GIS Working Group Workshop at the All Scientists Meeting in September, and will be distributed across the Network after that meeting.

The LTER sites have benefitted significantly from the Technological Supplement program over the past three years in the realm of Geographic Information **Systems**, remote sensing, local area networking, and to some degree, connectivity. The group consensus was to encourage NSF to continue the technological supplements with increased emphasis on connectivity. There are three sites (HER, CWT, LUQ) which are not connected to the Internet, which is an impediment to file transfer across the LTER network.

Another emphasis should be on personnel support as well as hardware and software acquisitions. Concern was expressed that full network connectivity will be difficult to achieve unless additional support is provided for technical specialists (i.e. UNIX technoweenies). For example, there was considerable interest expressed in archival of data on optical disks to enable long-term data protection and support decreased emphasis on inefficient and less secure forms of storage (i.e. tape storage). A further focus is in the area of database management. Additional funding will also be necessary to support the network global positioning system equipment (annual maintenance agreements, insurance,

upgrades) and support acquisition of more GPS units if warranted. The data managers would like to see the Technological Supplements Program continue to help the collection of LTER sites attain full network capability.

## F. DBMS/SQL

The subject of database systems for management of ecological data sets was a recurring **i~heme** at the meeting. In particular, questions relating to the conceptual type of database were discussed in detail (**11ierarchical**, flat, relational). Traditionally, most sites keep their archival data set copies in ASCII files, because this encoding is widespread, easily importable from and exportable to programs other than DBMS, and there is usually an easy way to read or modify files using only the most basic operating system commands Or functions.

Nevertheless, with increasing data volumes and a proliferating number of data sets, interest in more powerful data management tools is increasing. Relational database management systems **~BMS**) are now very common. Also, a query function for extracting subsets of large databases is an almost universal requirement. Similarly, indexing or sorting is often required in conjunction with the extraction of subsets. A number of sites use statistical packages, such as SAS (10 sites using SAS), to carry out these functions. This can be very convenient because it provides some degree of direct integration between data management and statistical analysis.

## 7

New methods of data management may be required, however, when access to **large** data sets is to be automated and data, or meta-data such as catalogs, are to be accessible over networks. One way of dealing with such a situation is based on the client-server model, a paradigm already common in software systems architecture. Database servers can be accessed over networks directly or by electronic mail query. Together with a query language standard, this mode of operation allows distributed databases to be built with heterogeneous hardware and software.

**Structured** Query Language (SQL) presently is the defacto industry standard for relational database query. It provides a high-level view of a database that is consistent with the relational model. The group concluded that it may become increasingly difficult for a site to justify not having a database system with SQL capabilities. (John Pfaltz offered to make shareware SQL available to interested sites.) An SQL-based system could be instrumental in building a network-wide distributed database capability, as would be needed in the climate database pilot project. An additional advantage of an SQL-based system would be that it could replace the INFO part of the ARC/INFO's GIS package with a much better user interface (ARC/INGRES, ARC/ORACLE).

It was emphasized, however, that the most reasonable approach is for a site to keep their present DBMS system, if that system provides the required functionality, and consider SQL capability as a high-priority function in any considerations for change. It was mentioned that most site PIs don't directly use DBMS software and therefore would not be directly affected by any system changes.

Several concerns about current SQL implementations were expressed and should be considered by those sites interested in exploring new DBMS software. First, SQL may solve only 5% of a site's problems. Secondly, the handling of text data is not strong in existing packages, but this is not a deficiency of SQL -- just of current implementations. Finally, the lack of a built-in **outerjoin** function in standard SQL is a serious impediment to using it for ad hoc merging of ecological data sets.

## G. STANDARDIZED DATA **DESCRIFION** FILES for INTERSITE DATA EXCHANGE

It was proposed that the use of standardized data description files be discussed at the All Scientists Meeting for consideration of adoption by all LThR sites. Two sites, SEV and CPR, are currently using machine readable data

descriptions to document flat ASCII files. The SEV system is modeled after that developed by Walt Conley, whereas the CPR system is an extension of a system used to document the Grassland Biome IBP data. Tom Kirchner agreed to distribute examples of the CPR data descriptions prior to the ASM.

A brief description of the two protocols follows: (1) Conley's **INTERSITH** protocol - a format for exchange of data which relies on ASCII files and uses simple

8

- tools for extracting data and documentation (including an abstract); and (2) a protocol which relies on ASCII files; descriptions are stored in machine readable form (including documentation). Labels for data fields, formats, units, and a brief description are included. Programs read this file to produce outputs and produce formatted descriptions. The **first** few records need to be in set format; after that, it is largely *free* form with associated labels.

## H. ARCHIVAL HARDWARE

William Michener (MN) and John Vande Castle reported that several options are available: CD-ROM (FIFE database); rewritable optical disks (Andrews, Network office, North Inlet). Several issues or concerns should be addressed when choosing this technology. The cost of media can vary from one machine to another. Frequently, it is desirable to purchase the optical disks from the manufacturer that sold the hardware unit. They can format the disk and test it in that specific drive. Costs for the rewritable drives may vary from approximately \$3,000 to 6,000 depending on the platform and available discounts. Media costs are approximately \$ 0.30 per mbyte vs. approximately \$ 2.00 per mbyte for floppy disks. Media life is approximately 10 years for optical disks vs. 1 year for magnetic media.

Back up units may be desirable, particularly when a **site's** database is being stored on optical disks. Data written onto an optical disk by a drive from a particular **firm** may not be accessible on drives built by other manufacturers. Multiple units may be viewed as feasible at a site when they are utilized as standard hard disks as well. Speed, therefore, can be a high priority consideration when choosing specific vendors.

Software varies according to manufacturer and several options are desirable. For example the ability to eject or format disks, etc.. may be important. The longevity of the various vendors should be considered (avoid fly by night companies). Alpatronix and Pinnacle were thought to be reputable companies with a proven track record in this field. Both companies have versions for most types of computer systems. Disks MAY be interchangeable between the same operating systems.

Pinnacle-15265 Alton Parkway, Irvine CA 92718 (800) 553-7070

Alpatronix-2300 Englert Drive PO 13687 Research Triangle Park NC 27709 (919) **544~o()()1.**

Both are rewritable using Sony drives and 600Mb (300/side - 1 side at a time) 5 12byte/sector disks for most machines. William Michener and Scott Chapal at NINLThR have been using the Alpatronix on a Sun. They report no problems: optical disks have been mounted, read from, written to and executed by SUN, PC's, and Macs (over TOPS). The Network Office has been using a Pinnacle on a Sun SPARCstation and 80386 PC under DOS. They have had problems using it under Interactive Unix on the **8038&PC.**

John Vande Castle had Alpatronics stop in for a demo to compare it to the Pinnacle. Both drives worked the same. They are fairly slow for writing and average disk access is about *65ms* - fast enough, but not fast enough for your main or only disk drive. The most important part - both systems can read each **other's** disk. He popped a disk written on the Pinnacle into the Aiphatronics, it worked fine, and so did the other way around. This is probably only true for similar file systems. This does mean that files could be sent across the network when needed.

Alpatronics does market software to be able to use disks from other file systems (like DOS or DEC-VMS) - This is still a to-be-released product (12/90) on the Sun but exists for the VAX. On the Sun, Alpatronics looks just like a standard **19sd** disk drive. They have special (copy protected) formatting and setup software. The Pinnacle has **it's** own drivers and formatting software. John Vande Castle has, however used the Pinnacle as an **"sd"** disk - but **doesn't** recommend it. (Both drives however, can be used as a boot device.)

One word of warning - 300Mb/side is not all that much with images and GIS files. Maxtor has a drive that is much faster (it uses a Tahiti subsystem) and can write non-standard larger format disks and perhaps (we have never used one) the standard ISO/ANSI 512 or 1024 byte/sector disks. Some sites (like NTL) have been using WORM drives like the 200mb IBM for the PC. These are also great, but once written, the files are unchangeable (they can be deleted, but the disk space remains used). Being able to move 200-300Mb of random access disk on and off systems is better than trying to use tape. For most purposes the disk can be used just a another hard disk. A word of warning however. One person in Forestry has an early version IBM WORM drive. It cannot read disk written on new versions of the same drive. Hopefully the ISO/ANSI standards have solved this type of problem, but I don't count on it. The Remote Sensing database from the satellite acquisition will be archived on the Pinnacle Optical and to-be-purchased secondary Alpatronics drive. We would appreciate further information/experiences or comments.

## I. DOE ECOLOGICAL RESEARCH PARKS

Dr. Walt Conley provided information on research activities at the DOE Ecological Research Parks. *Six* parks are currently in operation: Savannah River, Oak Ridge, Fermi/Argonne, Los Almos, Hanford, Idaho. A seventh site (Nevada test site) will be added soon. The parks are characterized by: large areas, few disturbed areas, high security, and large environmental research staffs.

Dr. Conley discussed characteristics of PARKNET which is similar to LThRNET and operated out of New Mexico State University. Additional discussion focussed on scientific workshops (how to get information on past workshops and plans for new workshops at NMSU) and intersite data exchange (see III.G.). Cooperation between LThR and DOE Research Parks regarding electronic communication, workshops, and

data management support groups received considerable attention. **PARK--T** now has a mailing group for their site data managers. You can send messages to this group by **sending** to:

**PARKdmg**~ lternet.washington.edu (or **PARKdmgr**~ lternet. bitnet)

## J. AGENDA FOR THE ALL SCIENTISTS MEETING

The following topics are currently on the agenda for the All Scientists Meeting.



## M. QUALITY ASSURANCE

Quality assurance has been, and continues to be, a hot topic for the LTER Data Managers. Although not discussed in great detail at the 1990 Data Managers Meeting, it was put on the agenda for the 1991 meeting. It will also be discussed at the May, 1991 Conference on Natural Resources Monitoring, to be held at Oregon State University (AND) next spring.

## N. DATABITS

John Porter (VCR) was commended for his efforts at making Databits such an overwhelming success. Future topics and distribution were discussed.

## O. tIdBITS

Barbara Benson (NTL) agreed to serve on the Data Management Task Force. James Brunt (**5EV**), William Michener (MN), John Porter (VCR), and Susan Stafford (AND), along with Rudolf Nottrott (NET) remained on the Task Force.

# IV. WORKING GROUP REPORTS

## A. CLIMATOLOGICAL DATABASES

Two models for climatological database development were proposed:

I. John Gorentz (KBS) proposed (together with one or two interested sites and the network office) a type of distributed climatological database with features as follows:

\* It would follow somewhat the pattern of a prototype developed at the Kellogg Biological Station (demonstrated at the April workshop) for distributed access to **herbariL**

~ Each site's climatological database would reside at the local site and be managed there.

\* The parameters being measured and the organization of the data need not be standardized from site to site, although standardization by mutual consent would be **encourageCL**

\* Electronic mail would be a primary transport mechanism. Participation in the database would be open to any site that has network gateways to the Internet. Rudolf Nottrott (NET) pointed out that those sites with a direct Internet connection should communicate

by more reliable means (such as remote procedure calls), even though they Communicate with systems on other networks by mail queries or definitions.

Each node that provides access to data would have an SQL interface to the data. SQL would be a standard query language, although at many sites the data user would not have to use SQL directly.

\* A standard format for transport of data, such as the Intersite Archive format (Conley (NMSU) & Brunt (SEV)) would be used.

\* The network office would participate, as one access point to the data, as well as to

provide any central information functions that are needed. The Network Support System proposed by the network office would help facilitate nation-wide access for users without any other network access.

\* The query mechanism would allow that potential data users to query database nodes for information about the form and contents of the data stored there.

It is assumed that all the databases at LTER sites, especially the biological databases, are and will continue to be a heterogeneous mix. An approach that allows for distributed access without enforcing unnecessary standardization is needed. It is hoped that this approach can be extended to other data beyond climatological data.

A few sites interested in developing a prototype, plus the network office, could develop a proposal to develop the necessary protocols and software. After using their own sites as a test bed, they could make it available to the others.

**II.** Tom Kirchner (CPR) proposed that we establish a prototype of a distributed database system using standard daily climatic data (minimum and maximum temperatures, precipitation, wind speed, **etc..**) for the database. The purpose of the prototype system is to demonstrate the feasibility and utility of a distributed database system. Cooperating

sites will provide one or more ASCII files of climatic data plus a data description file for each data file. The data description file is an ASCII file that describes the format of the data file, provides labels for the data, and gives other documentation of the data. Data description files are an important component of the data management system at the CPER **L~IER** site. The prototype of the distributed database system will be partially based upon software that was developed to access the CPER data via the data description files.

Those sites accessible via the Internet which have computers running BSD UNIX or SUN OS will provide an account on their computer that can be used to access the data. Kirchner will provide the software necessary to access the data across the Internet. The software will enable a user to display data from the remote site as a plot or a table. The

extraction and transfer of the data from the remote site will be handled entirely by the software system.

Those sites that do not have direct access to the Internet or are running non-UNIX proprietary systems will have the option of providing data files too be stored on computers at other sites. The feasibility of using electronic mail to request and deliver data and using PCs to display the data will be investigated if time permits.

## B. GIS/REMOTE SENSING/DATA MANAGEMENT SYMPOSWM

John Vande Castle (NET) led a small group discussion for and LTER GIS/Remote Sensing (RS) Data Management Symposium included a number of specific topics:

1. coordination with national committees
2. defining basic data layers everyone should have
3. standards of scale
4. nomenclatures and classifications
5. documentation to accompany GIS data layers
6. protocols
7. research aspects
8. preparation of report emphasizing standards

Such a symposium would need to include GIS, RS and data management groups within and outside of LTER. Especially important would be the expertise from other groups who already have database management for GIS type data in place. There is the need to recognize, however that the data involved in ecological research can be quite diverse resulting in a need for new standards. These standards would include, but not be limited to the type of data in GIS/RS data layers, standard scales, and format of the data. The types of data would include what minimum data layers are needed for ecological research such as Digital Elevation Models, soils, vegetation etc.. and the scales would include what scales of measurement need to be included in ecological research. Such a symposium would need to be cognizant of the future effort of NCGIA to integrate RS

data into "conventional" GIS data.

This symposium would be narrowly focused with a tangible product in the form of a report resulting from the symposium. This report would be entitled "**Suggested** Standards for Ecological GIS research". The specific topics of standardization of scales, basic data layers needed etc.. would have an invited speaker(s) to start discussion of each product. The initial framework. of this workshop would be part of the RS workshop of the LTER All Scientists Meeting to formulate a specific proposal. This proposal would fit very well into the new NSF initiative of "Database Activities in the Biological, Behavioral, and Social Sciences. John Vande Castle would coordinate the proposal writing with the help of people such as John Briggs, Skip Walker, Jim Halfpenny and

Others. The proposal would be sent to NSF by March 1 with an initial pre-proposal **resulting** from the **LTER** All Scientists Meeting.

## C. OUTREACH ACTIVITIES

Barbara Benson (NTL), John Porter (VCR), Eda Melendez (LUQ), Gil Calabria (CWT), Caroline Bledsoe (NSF), and

Susan Stafford (AND) are exploring offering a workshop on data management for Chinese LTER sites (CERN, the China Ecological Research Network). Our LThR experience could be shared with them and materials prepared for the workshop could be used to meet other requests for data management information both from **non-LTER** groups and from newer LThR data managers. Preparatory materials might include a collation of previous LThR data management documents and a short data management primer. The primer would raise issues and questions which should be addressed by people initiating a data management system and would direct readers to appropriate documents. It was suggested that there be a hands-on format to the workshop.

Members of CERN need to be approached to discuss their level of interest in such a workshop. The Andrews site already has strong research connections to the China group. A representative from CERN will attend the All Scientists' Meeting. Susan Stafford will pursue contacting CERN.

The advantages and disadvantages of extending the **L~I~R** mail forwarding system to the larger ecological community were discussed. The consensus was that the LTER mail forwarding system should serve as a model to construct additional systems which could be linked to LThRNET. To implement this larger mail forwarding system, it was suggested that Rudolf Nottrott collaborate with some interested non-LThR people in writing a proposal which would support a new person to set up and administer the new systems and the LTER system.

We want to make contact with other groups having data management expertise. Susan Stafford, Caroline Bledsoe and Barbara Benson agreed to research this issue and compile information on such expertise. We will then suggest ways in which the expertise could be made available to us.

A discussion ensued on helping data managers in their ongoing educational effort with their local PI's on data management issues. Many sites use regular oral or written reports on data management activities. John Porter agreed to make Databits available in ASCII form to facilitate extraction of articles for local use. Another suggestion was the development of a data management demonstration (perhaps using the network climate database) for use either at individual sites or at the Coordinating Committee meetings.

## 15

### D. PROPRIETARY ISSUES

LTER Data Managers recognized the complexity of this issue and felt that no definitive statement was possible without having input from LTER scientists, NSF, and **oher** interested parties. Nevertheless, pros and cons of data sharing, legal issues, and time limits for data "ownership" were discussed by the data management group as a precursor to the larger discussion planned for the All Scientists Meeting. A variety of points were made about each of the four items considered. They will serve as the basis for further discussion and are outlined below:

Potential positive aspects of data sharing and greater accessibility include:

- (1) Advancement of science - longer time series, broader spatial coverage, large scale synthesis.
- (2) Potential for enhanced collaboration among scientists and generation of new insights and ideas. Other investigators may see trends or results not apparent to the original PI.
- (3) New statistical and analytical techniques can be applied which the original investigator may not be

aware of or have access to.

(4) Greater attention by individual scientists to quality assurance and documentation may result. Furthermore, when data are shared among a variety of sites or researchers, the likelihood of data loss through local disasters (disk crashes, fires, floods, etc..) is greatly diminished.

(5) Accessibility to more data than any one individual can collect and ability to generate additional information based on corroborative data sets may result in many scientific advances. The time required by individual investigators to test hypotheses and the potential for duplication of efforts may be greatly reduced.

(6) LThR is based on the idea of "collaborative long-term **research.**"<sup>9</sup> The emphasis placed on data management by NSF and LTER is meant to insure that the data sets outlive the researcher.

Potential negative aspects of data sharing and greater accessibility include:

(1) Mandatory sharing may discourage participation on the part of some investigators. Investigators may feel that mandatory data sharing prior to publication is tantamount to theft of intellectual property. There is the additional concern that the shared data might not be acknowledged properly in publications.

(2) Money and time (additional work load) are required to document and manage data sets in a form that can be used by numerous individuals not connected with the original research.

(3) Data sharing could lead to misinterpretation of data: It is difficult to communicate information about idiosyncrasies and anomalies thoroughly enough, especially when the data being shared are in early stages. This can allow data to be used in inappropriate ways that eventually could reflect badly on the contributing researcher. New investigators may not have the necessary insights into specific data sets.

16

(4) There is the potential for spread of viruses, worms and bugs when data and programs are shared among numerous individuals.

(5) Some data sets gathered for a specific purpose may not be useful for future or **other**, as yet unknown, purposes. Extrapolation beyond the limits of a data set and inappropriate use of data for another project for which it was not designed could be a deterrent for data sharing. It may, therefore, be difficult to justify the additional expenses necessary to maintain selected data sets for long periods in a format that can be accessible and understood by other investigators.

(6) Potential legal problems (see additional discussion).

(7) Insecurity: initial investigator may have done research incorrectly or falsified some data.

(8) Tenure considerations - Currently few scientists receive credit for sharing their data.

Several legal issues must be considered by sites who are establishing guidelines for data accessibility:

(1) Violation of patents and copyright laws should be avoided. The question '**who** owns the data?' must be answered.

(2) University ownership of specific data sets (genes) and funding agency requirements.

(3) Potential use of data by businesses for resale.

(4) Local, state, Federal, and international laws protecting endangered or threatened

species may preclude data sharing.

(5) Misinterpretation of data sets in the courts by consultants or others may force the original investigator(s) to spend an inordinate amount of time correcting the mistakes or untruths perpetuated by others.

(6) Existing archival, quality assurance, and data integrity regulations or standards for specific types of data should be incorporated.

Data managers recognized the value of making data accessible to others in order to advance science. Many data sets are extremely valuable and could shed light on processes associated with global and regional change, biodiversity, and other timely scientific questions. It may, therefore be appropriate for sites to consider establishing time limits for data set accessibility. The following questions or factors should be considered during this process:

(1) Encouraging data accessibility after publication and completion of the scientific review process would ensure that "high quality" data sets are available for additional research purposes.

(2) Time limits for data accessibility may be different for monitoring vs. hypothesis oriented studies. The time period of the grant should be considered. Some research questions may require long-term data to answer.

(3) The inevitable death or **departi~** of key investigators at a site should be taken into account.

(4) A data set typically has NO VALUE until quality assurance and documentation standards have been met.

17

(5) One or more sites are considering establishing "3 years with appeal under special

**circumstances'<sup>t</sup>** or "released only following review for specified purpose" as their standards. The successes and problems associated with establishing these

standards have not yet been documented.

## E. META-DATA

Meta-data is that body of information needed to access, retrieve, and interpret scientific data. Individual site handling of meta-data vary greatly in structure and software used. Storage descriptions included ASCII, word processing (Word, XYWRITh), spreadsheet (Symphony), and database software (Foxpro, Ingress). File headers and comments within files are used at some sites. Front-end programs are also used to examine documentation or data. Certain qualifying parameters and historical methods logs are handled in different ways.

There is general agreement among sites about the types of information (metadata) included in the documentation files, but no standardization. It was generally agreed that standardization of software would be too hard to implement. Perhaps Conley's Intersite data exchange format could be adopted for intersite exchange of data.

Many sites seemed to agree that data set documentation data that only the **P1** could provide was difficult to obtain. What is the role of the data manager? It was noted that the explicit goal of LTER is the maintenance of long-term data and its documentation, and Prs should feel privileged to install data into a long-term database. The responsibility to provide this documentation should rest with the **P1**, and perhaps data should not be accepted until this information is provided. The Hubbard Brook sites demands an ASCII format file, quality control checks and documentation before accepting files.

Some suggestions to improve communication of documentation from the **P1's** included a) a meeting with the P1 in the planning process, and getting this information before data collection begins (new studies), b) providing more painless forms or data entry programs for entering documentation, c) have LTER data managers endorse a set of *minimum* documentation requirements that could be summarized in one page and distributed to Prs. Also, produce a more extended list of documentation requirements or options for use by **L~IER** data managers (especially new ones). The HJA will provide these products to the LThR data managers by email for review. d) criteria for site review should include documentation.

Other suggestions: a) include data documentation in P1's evaluation, b) rate data documentation in printed catalog and c) have the CC endorse a P1 requirement to provide documentation.

## 18

### Proposed Minimum Data Abstract Standards for Review

As promised at the LThR Data Managers meeting, the following letter to **LThR** P1's has been drafted for **everyone's** review. Don Henshaw, Susan Stafford, Gody Spycher (AND) are proposing that this document be distributed at the All Scientists Meeting in conjunction with the proprietary issues session, and considered for endorsement by all **LTBR** Data Managers.

*Title: Minimum Set of Meta-Data (Documentation) for all Data Abstracts of LTER Data Sets*

*LTER Data Managers, meeting in Snowbird, Utah, held a working group session on managing meta-data. Meta-data is that boay of **i~ormation** needed to **access, retrieve,** and interpret scientific data. After some review of how the various sites organized and maintained their meta-data, the discussion **shifted** to the role of the data manager in acquiring this documentation. Based on comments from the LTER Data Managers, one major problem is receiving necessary data abstract information which only the principal investigator can provide. A considerable quantity of information may need to accompany a given data set, but certain key information must be available before a data set can be properly installed for access in a long- term database.*

*Ideally, this information should be provided in the early planning process of a study. Unfortunately, investigators sometimes forget that including the Data Manager in planning discussions will improve data collection and processing. Furthermore, we are not suggesting that P1(s) involvement with data set documentation should end here, rather we recommend that the P! and Data Manager establish a long-term symbiotic relationship. From the Data Manager's perspective, it would be ideal if the principal investigators viewed installing data in a long-term database as highly desirable, and viewed cooperation with the site data manager on data documentation as critical to long-term database value.*

*We are proposing the following minimum set of standard information for data abstracts which the P1(s) must provide:*

*Study title: The title of the study.*

***~:** **Keywords** that will aid other researchers in review and retrieval of study data (a prototype list is available in the LTER Core Data Set Catalog for review by the PI(s) if necessary).*

***Parameters:** The primary variables that occur in the study.*

***Site location:** The specified study site location(s).*

19

***Study Purpose, goals:** A statement of the objectives and goals of the study as they relate to study data sets.*

***Experimental design:** An outline of the experimental or sampling design of the **5tU~** in sufficient detail to describe the basic experimental or sampling approach, plot size and **shape**, experimental unit(s), sampling unit(s), timing of sample, etc..*

***Methods:** A description of the method by which measurements were taken (both field and laboratory) with adequate detail provided to judge the propriety of potential comparison of data sets with regard to methodology.*

***Pro yrietary limits:** Specify access restrictions and an expected date when data become public property.*

*The above informational categories can initially be easily ignored by a data manager, as data set formatting and error checking are more immediate tasks. However, this documentation is essential in maintaining data set integrity; data managers must be vigilant in securing this information, and this commitment must be supported by the **PI's**.*

## F. NETWORKS and DATA MANAGEMENT

James Brunt (SEV), John Porter (VCR), Caroline Bledsoe (NSF), John Vande Castle (NET), Esteban Muldavin (JRN), and Emery Boose (HER) led a discussion on networking. An increasing number of LTER researchers make use of electronic networks in their efforts to keep up with the flood of information research resulting from ecological research. Most often, electronic networks are used **fo?** mail, individual or group mailings. This is in part due to historic reasons - in the past, most commonly available networks were designed with electronic mail as their primary function. The group estimated that the ratio of electronic mail traffic to traffic caused by other sources, such as file transfers or remote logins, is of the order of 10:1. The number of usernames in the LThRNET electronic mail forwarding system has increased from 150 in 1989 to 430 in 1990 and is expected to keep growing rapidly. Number and volume of file transfers are expected to increase considerably when the planned Network Support System has been installed and the **L~IER** network-wide remote sensing data acquisition has **materialize~** Also, publication of the LTER Catalog of Core Data Sets will stimulate action on data access issues. Pressure from groups within the larger ecological research community to make data accessible will increase and data already accessible will be used more frequently.

There is a need to extend networking capabilities as, they exist within the LThR network, to the larger ecological community, because networks provide the opportunity to do collaborative work more efficiently. The higher the number of ecological researchers already on an extended network, the greater the incentive for the "remaining" researchers to join. Beyond this effect of "**critical mass**", much needs to be done to make information retrieval over

the network easy and efficient. The on-line LTER Core Data

20

Set Catalog and the proposed distributed climate database provide opportunities to explore and develop user interfaces in the context of ecological research.

It was noted that networks provide an excellent medium to overcome problems of software and hardware compatibility. This will be **essential** in the implementation of future distributed databases.

Some concerns relating to present and future network use were discussed. One obstacle in the use of networks for data transfer is that transfer may still be too low for handling files of tens or hundreds of Megabytes in size (image files). This will in part be overcome when the NSFNET backbone is upgraded, as proposed. Also, Other concerns were discussed relating to the suggested use of database servers for making data or metadata accessible over networks; for example, derived data sets should not be accessible without accessing the corresponding meta-data. Another concern was about the possibility of privatization of the Internet several years from now. This could severely restrict scientific use of network.

Appendix I.

Attendance 1990 Data Managers Workshop

Name Site Name Site

Don Henshaw AND Cindy Veen **HBR**

Foresy Sciences Lab NE Experiment Station

3200 S.W. Jefferson Way U.S. Forest Service

Corvallis, OR 97331 P.O. Box 640

Durham, NH 03824

Gody Spycher AND

Forest Science Dept. Emery Boose HER

Oregon State University Harvard Forest

Peavy Flail 154 Harvard University

Corvallis, OR 97331-5705 Petersham, MA 01366

Susan Stafford AND Esteban H. Muldavin **JRN**

Forest Science Department Department of Biology

Oregon State University New Mexico State University

Peavy Hall 154 Las Cruces, NM 88003

Corvallis, OR 97331-5705

John Gorentz **KBS**

Bernie Moller ARC Michigan State University

Ecosystems Center Kellogg Biological Station

Marine Biological Lab 3700 E. Gull Lake Drive

Woods Hole, MA 02543 Hickory Corners, MI 49060-9516

Phyllis Adams **BNZ** John M. Briggs **KNZ**

University of Alaska Division of Biology

Institute No. Forestry Ackert Hall Kansas State University

308 Tanana Drive Manhattan, KS 66502

Fairbanks, AK 99775-0082

Eda C. **Melendez-Colom LUQ**

A. El-Haddi **CDR** Center for Energy &

**EBB/University** of Minnesota Environmental **Rese~h"**

318 Church Street, SE. Terresiriai Ecology

Minneapolis, MN 55456 GPO Box 3682

San Juan, PR 00936

**TomKirchner CPR**

NaL **Reso~ce** Ecology Lab John Vande Casde NET

Colorado State Uwvrsity University of Washington

Fort Collins, CO 80525 College of Forest Resources

Anderson Hall, **AR-IO**

**GilCalabria** CW-1- Seattle, WA 98195

School of Forest Resources

University of Georgia Rudolf Nottrott NET

Athens, GA 30602 University of Washington

College of Forest Resources

Anderson Hall, **AR-IO**

Seattle, WA 98195

**Name Site**

William Michener **~IN**

**Baruch Institute**

**University of South Carolina**

Columbia, SC 29208

Caroline **Bledsoe** NSF

Biotic Systems **and** Resources Division

National Science Foundation

1800 **G.** Street N.W.

Washington, D.C. 20550

Barbara Benson **Nm**

Center for Limnology

University of Wisconsin

680 N. Park Street

**Madison, WI** 53716

Jim Halfpenny **NwT**

**INSTAAR**

University of Colorado

Campus Box 450

**Boulder, CO 80309>450**

**Rick Ingersoll NWT**

**INSTAAR**

**University of Colorado**

**Campus Box 450**

Boulder, CO 80309-0450

James W. Brunt **SEY**

Department of Biology

University of New Mexico

Albuquerque, NM 87131

John Porter **VCR**

Environmental Science

University of Virginia

Clark Hall

Charlottesville, VA 22903

**Special Guest**

John Pfaltz

Institute for Parallel Computing

**Thornton Hall**

University of Virginia

**Charlottesville, VA 22903**

## Appendix II

### Site Summaries:

#### 1. Rudolf Nottrott (NET)

Rudolf Nottrott gave an overview of the present state of the Network Office computer systems.

Network Office hardware and software has been installed to directly assist and enhance research within and beyond LTER. Three heterogeneous, but integrated UNIX-based systems have been configured for LTER Network computer activities. All these systems share disks, tape and program storage resulting in a very powerful medium for information exchange. In addition, all three systems have been configured to implement the LTER mail forwarding system. A Sun SPARCstation-i has been installed to support network GIS and remote sensing activities. A VAXstation 2000 supports LTER's mail forwarding system, databases (Core data set catalog, personnel directory), bulletin board and wide-area network access. Both of these systems depend on an 80386-based PC-AT file server for data storage and peripherals. The file server's UNIX operating system also contains the VPIX environment which allows MS-DOS-based programs and files to be used within UNIX. This integrated computer system will allow access to large data sets including remote sensing image data across the Network. LTER Network-wide computer communications have been simplified with the University of Washington's implementation of Ethernet-based networking. The Network Office uses this framework as the basis for its local-area network (LAN), which is a subnet of the Internet (including NSFNET).

Following the report of the LTER Connectivity, the Network Office submitted a proposal to NSF

for a Network Support System (NSS). This initiative will implement the NSS on a network file server

that provides a computer account for researchers affiliated with the LTER Network, enabling them to receive or send Internet mail independent of their locations. By providing a centrally accessible access point to the Internet, the planned system will also support development of distributed network database

applications. The NSS will provide Internet services to LTER researchers who are located at unconnected sites, researchers on travel and researchers who are not able to get a computer account at their home institution. The NSS will facilitate file transfer (such as remotely-sensed images in the planned network acquisition of satellite data, manuscripts, software, etc.), remote logins to all computers on the Internet and access to on-line information on the Internet. The NSS will also be instrumental in efforts to expand the mail forwarding system to include service to the broader ecological community (e.g. Land-Margin Ecosystem Research sites, the Long-Term Study section of the ESA). The system will allow extending the capabilities with regard to connectivity, communication and database management, as they are presently available within the Long-Term Ecological Research Network (LTER), to the larger ecological research community.

**2. John Brigg (KNZ)**

Concerning our upgrade-- We got enough money to upgrade our SUN to put ERDAS 7.4 on iL In addition, we will purchase two 386 machines to run PC/ARC-INFO. Personnel wise, we also got enough money to keep our UNIX-Networking person for another 6 months. Other misc. items we will purchase, include some software for our Novell network.

**3. Tom Kirchner (CPR)**

Funding, BBS status, hard disk upgrades (Box Hill disks wiSyr warranty 600 MB \$2500) and collaboration proposals etc.. through CISE (Computing systems information and engineering) were discussed.

24

**4. El Haddi (C DR)**

Sun acquisitions, hard disk problems, and the Christmas office disaster were discussed.

**5. Jim Halfpenny**

Discovery of *an* email hole with 2-1/2 years of mail in it was found and fixed. GIS progress including: help from outside companies, problems with existing printers and the proposal for electrostatic printer was presented. The new Data Manager - Rick Ingersoll - was introduced to the group. Network extension to all users, climate database consolidation using Pascal **interface**, and use of matching funds for service contract were other topics.

**6. Emery Boose (liFR)**

Electronic mail through Omnet was implemented last December. In the coming year, with funds from LTER Supplemental Grants, we plan to set up a LAN connecting our research computers, and establish a link to the Internet, probably via a leased phone line to the Harvard campus network. We also hope to install PC ERDAS. Ongoing GIS work includes collaboration with Luquillo LTER to study the effects of hurricane Hugo, and a study of the relation between land use and forest vegetation in central New England.

**7. Barbara Benson ~TL)**

The North Temperate Lakes LTER was refunded and the 1990 technical supplement was funded. The supplement will be used to upgrade the LAN's at the Limnology Laboratory and the Trout Lake Station and to upgrade access to the databases. We considered upgrading the connection to the field station, Trout Lake Station, via a connection with **WISCNI-** but the cost was prohibitive, \$10K/year. The previous years supplements were used to acquire **015/Remote** Sensing workstations, create a **015** database, and connect the Limnology Laboratory to the campus Ethernet. Data management issues at our site which we hope will be discussed during this meeting include: database software on both the mini and micro's, mainframes versus micro's as a platform for data management, and storage media for long-term archival.

**& Cindy Veen (HBR)**

**Summary:** HBR had its site review this summer. There were few, if any questions about data management. A 015 Technical supplement for a PC ARC/INFO system for Hubbard Brook was submitted by cooperators located at Cornell University. The Hubbard Brook public bulletin board system 'The Source of the Brook' is up and running. Brochures for the bulletin board are available from the data manager. A building for the storage of permanent has been built at Hubbard Brook. A bar code system will be used as the permanent means of identifying samples.

#### 9. Eda Meimide: ~UQ

Routine business was discussed including: cataloging data (reference and LTER); DBMS (DBASE III+, now Paradox III on network LAN); getting documentation from researchers; established protocols; designing forms and full-time data entry personnel were discussed. Hurricane Hugo did not directly impact DM. Future considerations at LUQ include archival storage, Michener's consultation, upgrading computers, and online catalog in Paradox and graphics.

### 25

#### 10. Phyllis Adams (BNZ)

Progress on the fiber optic network; supplement - connect buildings to net & upgrade monitors; SUN 4 up and running in kitchen; ARC/INFO now loaded and running; and hiring a data manager 3/4-person were discussed.

#### 11. James Brunt (SEV)

The Sevilleta LTER program has expanded to include the Magdalena Mountains in the Cibola National Forest and the Bosque del Apache National Wildlife Refuge and USGS and NPS have been added to the list of agencies cooperating in work on the Sevilleta. All this has been done so far with no additional funding. The investigators on the Sevilleta LTER have been received a facilities grant to construct a real field station on the Sevilleta NWR. Two small grants have been funded to continue vegetation analyses from remotely sensed images in 2 and 3 dimensions. A proposal was submitted to NSF CISE for continued groupware visualization and graphics capabilities using the LTER network as a testbed, this was part of the "collaboratory" initiative. Networking for the new field station has now become a priority. The 1990 Science and Technology supplement was funded to provide a workstation version of ARC/INFO and to create Digital Elevation Models for the Sevilleta.

#### 12. Susan Stafford (AND)

- The Andrews Site has received encouraging words that we have been refunded for LTER3, although no definite word has been received as of this writing. The Technology Supplement has been funded, entitled; Technological Improvements to the Connectivity of the H.J. Andrews, GIS, and Data-tk". This will provide support to improve two aspects of connectivity at our site - hooking up the field station at the H.J. Andrews Experimental Forest with the University as well as connecting the Forest Service Data General System with our University LAN. We have also received a Facilities Grant to build a real field lab with housing accommodations at the Andrews. We have recently hired a GIS Geographer/Analyst (Sharon Clarke) with non-LTER funds. This will be the second GIS person added to the Quantitative Sciences Group personnel. Our GIS Support Programmer ~arbara Marks) was hired a year ago with seed money provided by the last Technological Supplement grant.

Last year's Data Managers Workshop report was well received at the Harvard Forest LTER/CC. Much appreciation is given to Caroline Bledsoe for orchestrating the agenda so as to provide an early, key time-

slot for our **repor~** This has facilitated a higher profile for data management activities throughout the past year. We, as **LTER** Data Managers, are enjoying much greater visibility than in the **pas~** We must now live up to the challenges this opportunity affords us]

There will be a Conference on Monitoring Natural Resources, May 6-9, 1991 Corvallis, Oregon. Susan Stafford and M McKee are the **co~ganizers** of this event. All LTER Data Managers are encouraged to participate. Anyone with specific ideas or suggestions, please contact Susan Stafford.

### 13. William Michener ~IN)

Hurricithe Hugo destroyed the Baruch Marine Lab on September 21, 1990 and much of the equipment was lost to the storm. Data was not lost, however, as it was archived in Columbia and other media was removed from the lab prior to the storm. Microcomputers and peripherals were also saved and are still in use. The field station is now operating out of the cottages and the Kimbel center, all of which have been networked together with a TOP's ethernet/Apple Talk system through a **FasLPath** gateway and a SUN Sparcstation has been added. Data management is being migrated from the mainframe to the SUN using the Alphantronix optical disk for data and program libraries to replace tape storage (much of the biological data is already duplicated on optical). A new data manager (Scott Chapal) is managing the biological data sets.

26

on the mam campus, **GIS** is now online with a PC dedicated to digitizing 8 hours/day and **~C~FO editing tasks** on a Sparcstation. Our technological supplement proposal was funded to expand the ERDAS installation and add scanning capabilities. The GPS proposal was also funded and **three** sites have volunteered to be regional centers. The **LTER** Coordinating Committee meeting (Puerto Rico) was summarized at the Data Management Meeting in Snowbird. Discussion **at** the LTBR-CC included the data catalog and proprietary rights issues.

### 14. John Porter (VCR)

Databits, PI change, remote sensing lab progress, networking, **GIS**, library acquisition of SPOT images, connectivity to shore, and EOSDIS were discussed.

### 15. John Gorentz (KBS)

**KBS recently put together the first draft of its data catalog, in preparation for the site review in**

early July.

In April a NSF-funded workshop on data management for field stations and coastal marine labs was held at **KBS**, sponsored by the Organization of Biological Field Stations and the Southern Association of Marine Labs. Three working groups, Data Administration, Data Standards for **Collaboraave research**, and **Computer Systems** for Data Management met and formed recommendations that will appear in a report to the National Science Foundation. The report will be ready for distribution to field stations and others this fall.

**LTER** was well represented at the workshop. Half of the rapporteurs who led the working groups are LTER data managers (Michener, Brunt, Porter). Other participants from **LTER** were Boose, Briggs, and **NOUtro&L**

### 16. Bernie Moller (ARC)

**Preparation for site review (postponed until next year) and the Sparcstation with ARC/INFO** up and going were discussed. E-mail and **FAX are working; no incoming** phone calls but can do 9outing.

**Main focus is on research through September.**

17. **Esteban Muldavin** (JRN)

PC Ingress as relational DM tool for users and a Sparc workstation that will act as server for data **sets were discussed.**

1& **Gil Calabria** (CWT)

**Networking, GIS activities, and recent progress in data management at Coweeta were discussed.**