

EXECUTIVE SUMMARY

The August 1989 LTER Data Managers' workshop focused on developing solutions and avenues of communication to solve commonly experienced problems and meet future technological challenges. The meeting succeeded in creating a broad consensus about what the major issues were and how solutions were to be pursued. An agenda (Appendix A) was employed which maximized group interaction and discussion. The Data Managers from all seventeen LTER sites attended and actively participated (Appendix B).

LTER DATA MANAGERS' RECOMMENDATIONS AND ACTIONS

Requests for Funding:

- * Develop an all-site LTER Catalog of key, long-term core data sets (PROPOSAL - Under separate cover)
- * Develop an all-site LTER Bibliographic Database (PROPOSAL-Appendix C-red tab) Decisions:
- * Establish a Data Managers' Task Force for 1 year to:
 - 1) Begin development of future priorities and goals for LTER data management
 - 2) Provide liaison with the LTER/CC by inviting a Task Force member to every LTER/CC meeting
 - 3) Facilitate exchange among the Data Managers' group
 - 4) Formulate, with input from the Data Managers, the agenda for 1990 meeting
- * Compile a set of LTER data management-related documents, to educate new data managers and increase continuity of the data management group
- * Establish a Data Managers' Newsletter, sent electronically and by mail, about every 2 months

Recommendations for Action and Discussion:

- * Establish research data management as a sixth LTER core element
- * Develop, in conjunction with the LTER GIS committee, documentation standards for GIS products
- * Develop a UNIX system administration workshop to facilitate rapid development of UNIX expertise at each site

TABLE OF CONTENTS

Page

EXECUTIVE SUMMARY i

TABLE OF CONTENTS ii

INTRODUCTION 1

REPORTS 1

NSF 1

Network Office 1

Network Connectivity Committee 2

WORKING GROUP SUMMARIES 3

Development of Dataset Catalogue 3

Development of Bibliographic Database of

LTER publications 3

Development of Task Force 4

Expectations of Data Managers 5

SPECIAL TOPICS 7

Optical Disks 7

SQL 7

Network Connectivity 8

Proposed Unix Workshop 8

Collaboration with GIS Committee on GIS

Documentation Standards 8

FUTURE ISSUES FOR 1990 MEETING AGENDA 9

SOFTWARE DEMONSTRATIONS 9

FISMENU 9

Andrews User Interface 10

SUMMARY 10

APPENDICES

A. Agenda 12

B. Participants 15

C. Proposal: All-site LTER Bibliographic database 17

D. Forest Science Data Bank at the H.J. Andrews LTER 21

INTRODUCTION

The August 1989 LTER Data Managers' Workshop focused on developing solutions to commonly experienced problems and on creating avenues of communication to facilitate solution of anticipated problems. The meeting succeeded in creating a broad consensus about what the major issues were and how solutions were to be pursued.

REPORTS

Caroline Bledsoe (NET/NSF) provided a review of the NSF administrative structure as it affects the LTER program.

Network Office

Rudolf Nottrott (NET) reported that the Network Office received a VAX station in April this year. In May, the VAX was connected to the national Internet via NorthWestNet, a regional network funded by the NSF and connected to the NSFnet high speed Internet backbone. The VAX'S Internet domain name is lternet.cfr.washington.edu, IP number is 128.208.36.1. Also, a modem connection was set up so users can connect to the LTERNET VAX over the telephone system, number (206) 543-2115.

During the wide-area networking workshop in Urbana-Champaign, it was suggested that a mail forwarding system be established for the participants in the LTER network. A mail forwarding system has been set up on the Network Office VAX and went through its first network-wide test during the preparations for the Toronto Data Managers' Workshop. It was used to send electronic mail to groups and individuals taking part in the workshop. Since then, all names and electronic mail addresses from the LTER personnel directory have been transferred to the mail forwarding database. (In addition, a text file containing all names, phone numbers and electronic mail addresses from the personnel directory is available on-line.)

The system greatly simplifies sending electronic mail messages by creating uniform addresses for everybody in the LTER network. An address to the forwarding system has the format username@lternet.cfr.washington.edu, whereby 'username' is formed for each individual according to a simple rule. A 'username' can also be group name. Mail addressed to a group name will be forwarded to all individuals on the group list. Presently there are three permanent group names: 'pi' for Principal Investigators, 'exec' for the executive committee, and 'datamgt' for Data Managers. Additional groups can be established at any time.

1

To form the 'username' for an individual, concatenate the first letter of the individual's legal first name and the last name. For example, James (Tom) Callahan's forwarding address is jcallahan@lternet.cfr.washington.edu ('tcallahan' will also work). The idea behind the scheme is that you shouldn't have to search through address lists - as long as you know somebody's name, you can always guess their forwarding address.

To further simplify Internet addressing, we are trying to shorten the LTERNET address to username@lternet.washington.edu.

In addition to the Internet address, the Network Office also has a Bitnet address. Bitnet mail to the forwarding system has the format username@lternet. Unfortunately, Bitnet imposes the limitation that usernames cannot be longer than eight characters. Therefore, to form a valid username for a Bitnet address, use the above rule and truncate the result to eight characters. Using the same example as above, Tom Callahan's Bitnet address would be jcallaha@lternet or tcallana@lternet.

As mentioned above, a text file containing some LTER personnel information is presently available on-line at the Network Office in Seattle, Washington. Rudolf plans to implement an interactive on-line directory providing the full information presently available in the LTER Network Personnel Directory. This capability, sometimes referred to as 'white pages', is part of the Andrew Message System developed with NSF funding at The Andrew Carnegie Mellon University and Michigan State University. Another function of the Andrew Message System is a multi-media bulletin board, capable of receiving and displaying graphics images.

Network Connectivity Committee

James Brunt (SEV) and John Porter (VCR) reported on the activities of the "Connectivity Team". Several of the questions being addressed by the Network Connectivity Committee were clarified. The charge of this group is to:

- 1) Assess the current state of connectivity to the National Internet
- 2) Develop long-term and short-term goals for network connections at all LTER sites
- 3) Develop plans and costs for implementing these goals
- 4) Report preliminary information to the NSF and the LTER Coordinating Committee (LTER/CC) at the October meeting and submit a final report by December 1, 1989

Data Managers were encouraged to meet with any of the Connectivity Committee members during the Data Managers' Workshop. Members of the "Connectivity Team" are John Porter, James Brunt and Rudolf Nottrott.

WORKING GROUP SUMMARIES

Development of Dataset Catalog

Several participants commented on the need for a database catalog indexing and describing the data sets at all LTER sites. It was also pointed out that the LTER Technology Report, presented by the Technology Committee at the April, 1989 LTER CC meeting in Albuquerque, New Mexico (SEV) lists the design and maintenance of database management systems for data and images as an important issue to be considered in the forthcoming LTER strategic plan. The discussion leader emphasized that control of research data at the site level would be preferable to a large centralized database, because this mode of operation transfers the question of access permission to the originators of the data and alleviates mass storage problems inherent in very large databases. Network access to the data sets distributed over the various LTER sites, as permitted by the site data administrator, should still be facilitated through a centrally located catalog and access mechanism.

The focus of the discussions on the subject of data and information sharing centered on the questions of what can be shared and exchanged on a regular basis, the scope of the sharing effort and suggestions for projects (including time frames and budgetary considerations for their implementation). The four working groups reached general agreement that the LTER Core Data Sets are sufficiently well defined and are common to all sites, and that it would be desirable and feasible to develop a catalog indexing and describing these data sets. This catalog should initially be available in printed form, but should ultimately be available on-line for interactive access. The catalog should build on the items proposed for the Network Data Set Directory in last year's summary of LTER Data Managers' Workshop (Michener and Stafford). A detailed project proposal for the implementation of the LTER Catalog of Core Data Sets is to be prepared by Bill Michener (NIN) and Rudolf Nottrott (NET) (under separate cover).

Development of Bibliographic Database of LTER Publications

The Data Managers discussed the development of an on-line LTER bibliography for all sites and the feasibility of undertaking this endeavor as a Data Managers' project. Topics covered included objectives and possible uses, what citations should be included, the possible format, and what steps should be taken to implement such a project.

3

Several purposes would be served by a network-wide LTER bibliography. First, Caroline Bledsoe pointed out the public relations value of this type of external document that emphasized the scope and breadth of the LTER network to the rest of the scientific community. A network bibliography would also provide public access to the literature of the LTER program and provide a database of research at all LTER sites to aid in intersite research efforts. Such a project would encourage the creation or updating of individual site bibliographies. And finally, such a network bibliography could serve as the initiation of an on-line computer network database of LTER-related publications with interactive access by all sites.

Two possible scenarios for an LTER bibliography are a complete version including all site-related publications, theses, etc. and a shorter version including only LTER-funded publications. The general feeling was that the shorter version would have limited utility and that the complete version, although it would require substantial funding, was the more desirable product.

It was the consensus of the group that a network bibliography should include a key word index if it could be created by one qualified person who could read every citation to reduce keyword variation among sites. A computerized version should have the capability to search titles, authors, and key words.

Clearly, the network bibliography envisioned by the Data Managers and PI's will require a professional librarian's expertise to compile. Individual sites would submit a bibliography with keywords. Data Managers expressed much concern that this task could not be done by the Data Managers and would require additional resources. Such a bibliography would need to be housed at a central location such as the Network Office and would need to be updated on a yearly basis.

The group decided to submit a proposal to the LTER/CC for funds to undertake this project (which will include funds for a professional librarian.) Phyllis Adams (BNZ) and Jim Halfpenny (NWT) will be responsible for drafting this proposal (Appendix C-red tab).

Development of Task Force

We, the Data Managers, decided that we did not want a steering committee. We want to maximize the coordination among the LTER Data Managers. We do not want one single individual speaking for the whole group. When requests come in to anyone, that person should immediately forward the request to the entire network and solicit input from the greater group. We agreed to form a Task Force with the following four functions:

4

- 1) To formulate, with input from the Data Managers, the

agenda for 1990 meeting

2) To provide liaison with the LTER/CC

3) To begin development of future priorities and goals for

LTER data management

4) To facilitate exchange among the Data Managers' group

Selected to serve on the Task Force until the 1990 Data Managers'

Workshop were: Susan Stafford (AND), Bill Michener (NIN), James

Brunt (SEV) and John Porter (VCR).

Expectations of Data Managers

A memo was sent from Jerry Franklin (NET) to all Principal Investigators of record directing them to discuss their expectations of data management with their data manager prior to this meeting. Many did, however, not all.

Participants were asked to synthesize those discussions and assess them with regard to their own expectations and their perception of the expectations of the NSF. Four working groups were formed with group leaders assigned the responsibility to report back to the main group-as-a-whole with the consensus of the subgroup. Out of these discussions there were three recommendations extended for consideration by the LTER/CC.

1) Data management be recommended for elevation to a core area of LTER. This in light of the critical importance of data management to the overall goals of long-term ecological work and the increasing importance data management plays to intersite work. As more and more data are incorporated into databases there will be increasing demands placed on data management staff. These increases are already being felt at older sites.

2) Discussion and synthesis of the criteria by which data management should be evaluated in the NSF review process be included as an agenda item for the 1990 Data Managers' Workshop. PI's expect data managers to identify and act on these review criteria. The Data Managers should be in a position to make recommendations to the LTER/CC directed at establishing these criteria.

3) A member of the Data Management Task Force be included on the LTER/CC to provide a liaison to the group whose charge would be to provide realistic expectations of data management and represent the collective group of Data Managers. The Data Managers

5

could better address issues of importance to network.

Discussion Notes:

The four working groups returned similar syntheses of expectations. The strongest responses are summed up in following paraphrase:

"PI's and the NSF look to the Data Managers to set their expectations...they look to Data Managers to anticipate future needs and facilitate research."

Three recommendations were made as an effort to alleviate some of these problems, more accurately reflect the importance of research data management, and elevate the status of data management within the network.

1) Archival and accessibility (both short and long term), data integrity, quality assurance and documentation were considered the primary functional expectations. There was a strong concern that although these perceived "lower-level" functions were rated highly important, there are unrealistic views among PI's of the person time commitment necessary to achieve these tasks adequately. It was noted that these functions must be supported and in place before functions can occur that will increase productivity and intersite comparison.

2) Other expectations returned by the working groups included assessing data for intersite comparison, providing summary analyses and graphics, advising on statistics and technology, along with expectations that data management at their site would increase the productivity of researchers and the quality of research. With these high expectations placed on the Data Managers, the group felt that data management did not have adequate status with the PI's and the LTER/CC. The more experienced site Data Managers encouraged newer sites to begin orienting the PI's to the idea of a "data management staff" instead of data manager, deeming it would be necessary to have more personnel dedicated to data

management to even modestly meet these expectations.

3) In the introduction of this discussion it was expected that there would be a broad difference between the groups' summary expectations and their views of what the NSF's expectations were. Contrary to this, it was the feeling of most of the groups that the NSF's expectations of data management may be the same as the

6

PI's and the Data Managers except at a network scale of operation and that better communication is needed among these three groups, i.e., the NSF, PI's, and Data Managers. Several sites indicated that they did not think that the NSF's expectations were being addressed in reviews of their sites, and that high-tech demonstrations carried more weight with the reviewers. The Data Managers strongly felt that the panel of reviewers should include a person highly familiar with data management and educated in the review criteria.

SPECIAL TOPICS

There were several topics discussed during the special interests session:

Optical Disks

Everyone embraces the idea of secure archival of data. WORM media offer a potential solution to the archival problem. In addition, optical drives may be preferable to tape for working with large data files such as those generated from satellite imagery because of access speed. However, there is no standard, costs are high, and the technology is rapidly changing. The recommendations are:

- 1) It is not appropriate to force standardization among sites.
- 2) Those sites that need a mechanism to archive large data sets may purchase a WORM to satisfy their needs but should be prepared to buy new machines in the near future as standards emerge and the technology stabilizes.

SQL

The discussion ranged over what SQL (Structured Query Language) is and how it relates to database systems. General conclusions were:

- 1) No outstanding database software currently exists.
- 2) Most sites store data as ASCII flat files and documentation is usually on a database system. The transportability of ASCII files allows investigators to import them into other systems for analysis.
- 3) It was suggested that individuals should read the discussion of database software inadequacies in the "Forest Science Data Bank at the H. J. Andrews LTER"

7

document distributed at the meeting (Appendix E-green tab)

- 4) No specific recommendations for SQL were given time.

Network Connectivity

Concerns about network connectivity were discussed. It was expressed that the site specific nature of this venture be appreciated. Concern was also voiced about the reluctance among some PI's to be connected to the national network because of security and personal reasons. On the other hand, concern was expressed that current email systems were not easy enough for some researchers. Although further discussion emphasized that "difficulty of the system" was merely an excuse. If PI's and other researchers want to use the email system, they will. It was suggested that if the NSF began sending out renewal notices via email, everyone would soon learn to use it!

Proposed UNIX Workshop

Tom Kirchner (CPR) raised the question of whether a workshop dealing with UNIX system management would be of benefit since many sites are acquiring SUN machines. The Natural Resource Ecology Lab has a staff that is capable of presenting such a workshop. The brief discussion about this question showed strong support for such a workshop. However, it was also suggested that, to be of greatest benefit, the workshop should be held in the near future. Tom Kirchner agreed to pursue setting up such a workshop at Colorado State University.

Collaboration with GIS Committee on GIS Documentation Standards Potential problems with integrating GIS and traditional data management tasks were discussed. Specifically, concerns over documentation standards, quality assurance, and data transfer were expressed. Six to ten Data Managers indicated that they would be attending the GIS workshop. A decision was made to have those Data Managers attending the GIS workshop, meet with members of the GIS committee to develop a proposal to resolve these specific concerns.

Two software programs were briefly discussed:

- 1) ARC/DYNAMO for ecological modeling
- 2) An ARC/INFO - SAS interface program developed by Oak Ridge National Laboratory

More information on these packages will be forwarded to the LTER sites as it becomes available.

8

FUTURE ISSUES

3 Both short term and long-term issues were identified. Short-term issues, to be resolved in the next year included establishment of an LTER-wide data management newsletter, creation of an online LTER personnel directory at the Network Office, and recognition of research data management as a sixth LTER core element.

A preliminary agenda for the 1990 Data Managers' Workshop was discussed and included the following items:

- 1) Criteria for evaluating LTER data management
- 2) The Kirchner Principle; "Heading off data before the pass" or "The use of machine readable headers for automated processing of archived datafiles"
- 3) Facilitating the research process at the site and network level
- 4) Review of 1989 Data Managers' Workshop
- 5) Problems with database management of models

Long-term issues and activities included a survey of datasets applicable to studies of global change, implementation of full connectivity of LTER sites across the electronic Internet, investigation of possibilities for joint LTER/IGBP workshops to address data management issues for wide-area data, exploration of linkages between GIS Systems and conventional relational databases and statistical programs, obtaining global georeferenced coordinates for existing datasets at all sites, identification of new technological resources with application to data management, and development of standards for machine readable data descriptions.

SOFTWARE DEMONSTRATIONS

FISMENU

The FISMENU program allows an investigator(s) to create command files to access the FIFE Information System (FIS) without knowledge of SQL, (Structured Query Language). By running the program on a PC, a person can actually write a program to access the FIS which is set up using the Oracle SQL subset. The program has three main screens; the Menu screen for viewing the menus and choosing an Oracle data table, the Format building screen for selecting columns from the chosen table, and the Query building Screen for specifying a limited range of data. (For example choosing the LTER species composition data for March 1987 as opposed to choosing the entire species composition for all the

9

FIFE site). Once the query is completed, the investigator(s) can upload the "program" to the FIS and actually get the data requested. The actual steps, of course, depend on your telecommunications limitations. This type of system may be applicable to the 'future' LTER data system.

Andrews User Interface

The H.J. Andrews User Interface Demonstration presented an early version database management program being written by the Andrews LTER data management staff. This menu-driven interface is written in the DBMS language Foxbase (dBase family), and is a useful tool for integrating data catalogs, documentation, and data files. The program is designed to allow direct investigator viewing and searching of data catalogs and documentation. It will also provide a means of archiving and retrieving data, including utilities interfacing datasets to SAS and other analysis systems. A Report Generator and User Help Screens are planned.

The demonstration showed how the Andrews LTER and the Forest Science Data Bank (FSDB) study catalog could be searched by keyword, category code, project code, or principle investigator. For example, all study titles for a given PI could be displayed, or all dataset codes referencing a specific keyword could be listed. A dataset documentation entry/viewing subroutine was also demonstrated. More information on the Andrews LTER Data Bank and User Interface can be found in the handout distributed in Toronto (Appendix D).

SUMMARY

The LTER Data Managers had an excellent meeting in Toronto, Canada, August 4-6, 1989 in conjunction with the AIBS/ESA meetings. Final recommendations and actions fall into three categories and are listed below:

Requests for Funding: -

- * Development of an all-site LTER Catalog of key, long-term core data sets (PROPOSAL - Under separate cover).
- * Development of an all-site LTER Bibliographic Database (PROPOSAL - APPENDIX C - Red tab).

Decisions:

* Establishment of a Data Managers' Task Force for 1 year to:

1) Formulate, with input from the Data Managers, the agenda

for 1990 meeting

10

2) Provide liaison with the LTER/CC by inviting a Force member to every LTER/CC meeting

3) Begin development of future priorities and goals for LTER data management

4) Facilitate exchange among the Data Managers' group

- * Compilation of a set of LTER data management-related documents, to educate new Data Managers and increase continuity of the data management group
- * Establishment of a Data Managers' Newsletter, sent electronically and by mail, about every 2 months

Recommendations for Discussion and Development:

- * Establish research data management as a sixth LTER core element
- * Development, in conjunction with the LTER GIS committee, of documentation standards for GIS products
- * Development of an UNIX system administration workshop to facilitate rapid development of UNIX expertise at each site

This report will be presented and discussed with the LTER/Coordinating Committee at the October 1989 meeting, Harvard Forest. The 1990 meeting of the LTER Data Managers will be July 27-29, 1990, Snowbird, Utah preceding the annual ESA meeting.

APPENDIX A

Agenda: LTER Data Managers' Workshop

Toronto, Canada, August 4-6, 1989

FRIDAY, AUGUST 4

Dinner on your own. Social, 7:00 - 9:00 pm (room to be posted), Chestnut Park Hotel, 108 Chestnut Street, Toronto, Ontario M5G 1R3, phone 416-977-5000.

SATURDAY, AUGUST 5

All meetings at Hotel, room to be posted. Breakfast on your own; coffee/tea/pastries available, 8:00 am, meeting room.

8:00-10:00 Introduction: Stafford (AND) & Michener (NIN) Updates on sites: all 17 representatives report briefly, about 3 minutes each!

Report on Connectivity: Brunt (SEV)

LTER/CC & NSF activities: Bledsoe (NET/NSF)

Network office activities, including LTERNET mail-forwarding system & naming rules: Nottrott (NET)

10:00-10:20 Coffee/tea. Demonstrations of mail-forwarding

system (Nottrott) and "User Interface" (Stafford).

Display of site bibliographies (please bring a copy from your site for display & donation to

LTERNET).

10:20-11:50 Discussion #1: What are the NSF's expectations for data management? What are the PI's expectations? Can these expectations be met?

Leaders: Brunt & Bledsoe

Discussions arranged as follows:

Overview
by leader
10-15
min

Small
group
discussions
40-50
mm

Large
group
exchange
&
summary
20-30
mm

11:50-1:00 Lunch (in meeting room) and free time

1:00-2:30 Discussion #2: Development of an on-line LTER Bibliography for all sites: A prospective data managers' project?

Leader: Adams (BNZ)

2:30-2:50 Coffee/tea

12

SATURDAY, AUGUST 5

2:50-4:15 Discussion #3: Sharing Information and Databases:

A critical component of ecological research

Leader: Nottrott

4:15-5:00 Discussion #4: Special Interest Groups

Leader: Benson (NTL)

Participants suggest discussion topics (e.g., relational database software and workstations, data retrieval, optical disc storage, SUN SPARC).

NOTE: Discussions will be aided by a no-host bar/snacks!

5:00-7:00 Dinner together (to be arranged by Stafford)

SUNDAY, AUGUST 6

Breakfast on your own; coffee/tea/pastries available, 8:00

am

8:30-9:15 Discussion, review, and synthesis of group decisions on Topics 1, 2, 3, and 4.

Identification of writing assignments for summaries.

Leaders: Adams, Benson, Brunt, Nottrott

9:15-10:30 Discussion #5: How will GIS integrate with data management over the 3-5 years?

Leader: Michener (NIN)

10:30-10:45 Coffee/tea

10:45-11:15 Organization of the Data Managers' Committee: Do we need a steering group? What are our objectives?

Leader: Stafford

11:15-12:15 Discussion #6: Future projects and activities for the Data Managers' Committee

Leader: Porter (VCR)

12:15-2:00 Working lunch and preparation of summary reports for Discussions 1-6

2:00-3:00 Wrap-up session

Leaders: Stafford & Michener

Recommendations to the LTER/CC. Agreement on projects to be initiated and assignment of responsibilities. Planning for next meeting (agenda, focus, selection of planning group).

13

SUNDAY, AUGUST 6

Distribution of summary reports to group. Using these reports, Stafford will report to the LTER/CC in October 1989.

3:00 Adjourn, see you next year!

NOTE: The next Data Managers' workshop (July 28-29, 1990) preceeds the annual

ESA meetings at Snowbird, Utah.

NOTE: The 1989 steering Committee (Stafford & Michener [chairs], Adams, Bledsoe, Brunt, Nottrott and Porter) developed this agenda.

APPENDIX B

LTERR Data Managers' Workshop Participants

Toronto, Canada, August 4-6, 1989

Susan Stafford (AND) Emery Boose (HFR)

Forest Science Department Harvard Forest

Oregon State University Harvard University

Peavy Hall 154 Petersham, MA 01388

Corvallis, OR 97331-5705

David Lightfoot (JRN)

Don Henshaw (AND) Dept. of Biology/SAF

Forestry Sciences Lab New Mexico State University

3200 S.W. Jefferson Way Las Cruces, NM 88003

Corvallis, OR 97331

Stephan Ozminski (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

Bruce Haines (CWT)

School of Forest Resources John Briggs (KNZ)

University of Georgia Division of Biology

Athens, GA 30602 Ackert Hall

(8/6/89 only) Kansas State University

Manhattan, KS 66502

Phyllis Adams (BNZ)

University of Alaska Eda Melendez-Colom (LUQ)

Institute No. Forestry Center for Energy &

308 Tanana Drive Environmental Research/

Fairbanks, AK 99775-0082 Terrestrial Ecology

GPO Box 3682

A. El Haddi (CDR) San Juan, PR 00936

EBB/University of Minnesota

318 Church Street, S.E. William Michener (NIN)

Minneapolis, MN 55456 Baruch Institute

University of So. Carolina

Tom Kirchner (CPR) Columbia, SC 29208

Nat. Resource Ecology Lab

Colorado State University Barbara Benson (NTL)

Fort Collins, CO 80525 Center for Limnology

University of Wisconsin

Cindy Veen (HBR) 680 N. Park Street

NE Experiment Station Madison, WI 53716

U.S. Forest Service

P.O. Box 640

Durham, NH 03824

15

Jim Halfpenny (NWT)

INSTAAR

University of Colorado

Campus Box 450

Boulder, CO 80309-0450

James Brunt (SEV)

Department of Biology

University of New Mexico

Albuquerque, NM 87131

John Porter (VCR)

Environmental Science

University of Virginia

Clark Hall

Charlottesville, VA 22903

Rudolf Nottrott (NET)

University of Washington

College of Forest Resources

Anderson Hall, AR-10

Seattle, WA 98195

Caroline Bledsoe (NET/NSF) Biotic Systems and Resources

University of Washington Division

College of Forest Resources National Science Foundation

Anderson Hall, AR-10 1800 G Street N.W.

Seattle, WA 98195 Washington, D.C. 20550

APPENDIX C

Proposal Submitted to the Long-Term Ecological Research Coordinating Committee

Title: Long-Term Ecological Research Program (LTER) Network Bibliography An Intersite Project

Authors: Phyllis Adams, Bonanza Creek Experimental Forest James Halfpenny, Niwot Ridge Alpine

Submitted by: The LTER Data Managers

PROPOSAL FOR DEVELOPMENT

LTER Network Bibliography

The Data Managers propose the development of a network-wide LTER bibliography. Access to a complete LTER bibliography is necessary for optimal LTER Networking capability. This bibliography would serve five purposes:

Enhance visibility of LTER network by:

- 1) Creating an external document emphasizing the scope and breadth of LTER science to the outside research community
- 2) Providing public access to the literature of the LTER program
- 3) Providing a database of research at all LTER sites for the enhancement of intersite research
- 4) Providing a computer network access to LTER research literature
- 5) Encouraging the creation and/or updating of individual site bibliographies for use by local investigators for research and publication

The Network bibliography would be composed of separate bibliographies developed by each LTER site.

The Data Managers were surveyed at their 1989 annual meeting to determine the importance of this project, its magnitude, procedures for development, and time frames and costs. There was agreement on the need for the development of a network

bibliography. Currently, LTER sites report over 1,400 funded publications and over 10,000 site publications. About 50 percent of the sites have developed key word indices. However it was felt that a quality bibliography that would be available to all audiences could only be developed under the direction of a professional librarian. This would assure maximum usefulness by providing a finished product consistent with

the national library base. The development of the bibliography would take from six months to two years and the total cost would be from \$11,000 to \$33,000 depending on the scope of literature cited for each LTER program and research site. This bibliography would be compiled from the individual site bibliographies of each site.

Two possible strategies for the bibliography were developed. Plan A would be a listing of all publications pertaining to each site. Plan B includes a listing of only LTER-funded citations.

Both Plan A and Plan B contain some of the same developmental steps. First, requests for proposals would be issued to selected personnel skilled in library sciences and bibliographic database development. The scope of the contents of the bibliography (refereed journal articles, theses, abstracts, reports, and/or pertinent literature) would be determined by a consensus of scientists at each site in consultation with the librarian. Selection of a bibliographic database and formatting would also be under the guidance of the librarian. The librarian could develop a finalized key word index and be assisted by the LTER Network office in providing interactive computer access. If a keyword index system is used, it is imperative that there be one qualified person who will read all the articles and assign keywords. Otherwise there will be too much inconsistency in keyword assignment from site to site. An alternative, would be not to use a keyword system and have potential users search the title for certain words. This would be discussed more fully once a decision is made to go ahead with the project.

Plan A

Plan A would provide complete access to the research literature of each LTER site including historical references. The bibliography would contain over 10,000 citations and would take one to two years to complete. Each site at this level would provide a complete site bibliography to the librarian for key wording. Currently only half the sites have undertaken the development of key word indices. Development and completion of key word indices for the rest of the network might add considerable time to the completion of a compiled bibliography. Input would be in the form of computer readable ASCII files and hard copy. The librarian would complete the Network Bibliography.

We have estimated that the development of a complete bibliography would cost \$33,000 as follows:

18

Librarian (50% time, 1 year) \$16,000

Computer 5,000

Typist/key entry 7,000

Bibliographic software/networking 5,000

A computer would be needed for one year to develop the bibliography, and could be used as a network server and for further update of the bibliography. Typist help is necessary to edit and/or enter citations. Bibliographic software that is of library quality and compatible with national library services would need to be purchased for the network office. This bibliography should also be made available to other library listing services. An alternative to computer and software purchase, although less desirable because of no on-line accessibility, is rental of an existing system.

Plan B

Plan B development would provide quick access to a completed bibliography for the network for only LTER-funded citations. This bibliography could possibly be completed in as few as six months, depending upon the state of completion of individual site bibliographies. Approximately 1,400 bibliographic citations would be included. Each site would provide a bibliography for their site. Each bibliography would include key words developed in consultation with the librarian. Again the input would be in the form of computer readable ASCII files and a printed copy. The librarian would complete the Network Bibliography.

We are requesting \$11,000 in order to complete this level of development as follows:

Librarian (50% time, 2 months) \$6,000

Typist/data entry (50% time, 2 months) 2,000

Computer time on existing software 3,000

However, if a computer and existing software are available for use, perhaps donated by a site, costs might be as low as \$8,000.

Bibliographic Updates

The Network Bibliography will have to be continually updated on an annual basis. Once established, the updating process should be a secretarial duty. We are requesting one month secretarial salary be allocated to the Network Office to provide this service each year. Each site would be responsible for sending yearly written updates.

Networking

For optimum access the Network Bibliography must be maintained on an interactive computer network. Development of this phase

should be coordinated through the Network Office. One possibility is to maintain the bibliography on an existing site computer. However, it is desirable to locate the bibliography in the Network Office for ease of update and maintenance. Therefore, it is suggested that a separate computer server be purchased and located in the Network Office.

Access to the Network Bibliography

In addition to on-line interactive access to the Network Bibliography, copies can be provided in an ASCII format to each site for use in their specific databases. Copies of the ASCII format would be available on-line or in disk format. Other inquiries may be answered in either format.

It may be desirable for each site to obtain duplicate copies of the bibliographic software for use at their site. This additional cost is not covered in this proposal.

APPENDIX D

Oregon

Department of **State** Peavy Hall 154

Forest Science **University** Corvallis, Oregon 97331-5705 (503)754-2244

The Forest Science Data Bank at the H.J. Andrews LTER

Introduction

The recent establishment of a networked-PC environment has enabled us, for the first time, to begin building an integrated data bank with support for a critical set of data base management functions. Guided by the recommendations of the Kellogg workshop (National Science Foundation, 1984), we have considered solutions which would provide adequate user support while minimizing the need for in-house programming. This summary describes the design and current state of the Forest Science Data Bank (FSDB).

Storage Structure

The design of storage structures for documentation and data on the LAN-servers is straightforward. The entire databank resides on one of our file servers (namely FSDB) under the FSDB subdirectory. For each study, data and documentation for each study are stored in a subdirectory whose name is identical to the unique study code (Figure 1). Catalogs and interface programs are located in separate directories. Data are stored as flat ASCII files, all catalogs and documentation files are data base management system (DBMS) files. Tape cartridges are currently our primary backup system. We are exploring the possibility of attaching a WORM drive to our network for secondary backup storage.

Catalogs

Currently, seven catalogs exist. The field lists and keys for each catalog are evolving and follow closely the Kellogg recommendations (See Appendix I).

The 3 main catalogs are:

- Studies - Lists all studies and associated information
 - Formats - Lists all formats and associated information
 - Datafiles - Lists all data files and associated information
- The supporting/accessory catalogs are:

FIGURE 1: FOREST SCIENCE DATA BANK

Server:

FSDB

SAS

FSDB

TVO10 Study

TVO10.ABS ABSTRACT **Subdirectories**

TVO10.COD CODES

TVO101.FMT

- FORMAT

- FILES

TVO106.FMT

TVO101.DAT

- DATA

- FILES

TVO106.DAT

PRG

FS.FOX

CATALOGS

STUDIES.CAT PI.CAT

FORMATS.CAT PROJECTS.CAT

DATAFILES.CAT CATEGORY.CAT

KEYWORDS.CAT

INDICES

22

Projects - Lists research projects (with funding source -both public and private)

Categories - Lists broad study categories; is also used to assign new study codes

PI - Lists principal investigators

Keywords - Lists keywords derived from abstracts, study titles, and dataset titles

The central catalog is the Studies catalog and it is "keyed" to the Projects, Categories and PI catalogs. Studies, Formats and Datafiles are three non-redundant hierarchical catalogs. Formats is keyed to Studies and similarly, Datafiles is keyed to both Formats and Studies. Keywords is a separate catalogue for study searches by keywords.

A catalog that relates geographic locations to studies is planned. A publications database, recommended in the Kellogg Report, was available in the old system, but was very difficult to maintain. We now have a departmental bibliography on PROCITE and are exploring ways to make this a searchable network database.

Data Documentation

The documentation files for a given study include several files which reside in the study subdirectory (see Figure 1). They include one file for each format, one file for variable codes, and one abstract file (see Appendix I for fieldlists). The fieldlists of formats and codes are very similar to those in the old system. Abstract files have been reorganized to eliminate information redundancies. We have opted to use ASCII text files covering the usual abstract topics (study purpose, locations, study design, experimental methods, etc). Abstracts, viewed on screen or written to printable ASCII files, also include pertinent information from the catalogs.

Function

General users have read-only access within the FSDB. Active users may copy files (catalogs, documentation, data) or read files through programs. New or revised documentation or data must be submitted to the data manager for quality assurance and accuracy checks before they become established in the data bank.

A User Interface provides access to the catalogs, documentation, and data. It supports entry of documentation and data by users, and facilitates export of data into a variety of analysis systems (e.g. SAS, Cornell Ecological Programs, etc).

23

User Interface

Need: Aiding research productivity through data management

has been frequently emphasized (National Science Foundation, 1984, Stafford et al., 1986, 1988). Our bare-bones, read-only data bank, as previously existed on the mainframe, did little to further that goal. Utilities which perform or aid basic data base management system (DBMS) activities needed to be provided for the benefit of users and the data manager (DM). Users should be able to find, document, enter, check, and export data with little or no assistance from the data manager. Therefore, we needed a User Interface - the development of which is a non-trivial software project of moderate size and scope.

Software choices: Establishing a user interface is closely linked to software choices for data management. We have been warned against reinventing the wheel and urged to use the capabilities of DBMS's (National Science Foundation, 1984). Unfortunately, software houses have offered little to support the scientific research environment where multiple independent databases and variety of analysis systems are the rule.

The newest versions of several successful DBMS's with full system query language (SQL) are useless for our purposes because they lack functions that operate across databases. Importing a new table into an existing database is typically not a programmable operation and in most cases, only one database can be opened at a time. These packages are intended mainly for business environments where only a few, static databases are used. In our research we are continually updating and adding new databases and have as many databases as we have study codes (circa 400).

The "ideal" DBMS for our environment should 1) permit simultaneous access to tables from several separate databases, 2) contain an adequate set of commands to manipulate and transfer data structures, and 3) offer flexible windowing- and text-writing facilities. We have concluded that the Dbase family (specifically pseudo-compiled Foxbase), in spite of certain shortcomings, provides the flexibility and speed that we need. We have also acquired a full-fledged SQL system (Rbase For DOS) for use in specialized situations.

Approach: Currently we see no need to alter the arrangement used in the previous version of the databank where data were stored as flat ASCII files for easy access by analysis systems, and both catalogs and documentation files were kept as DBMS files. The menu-driven interface written in the DBMS language (i.e. Foxbase) integrates catalogs, documentation and

data in ways that relieve users of certain tedious chores.

24

Users are not required to know the DBMS language. The main disadvantage of this approach is that SQL-type operations on data (i.e. merges, intersects, etc) cannot be provided. And accomplishing these tasks requires pre-analysis processing. At our site, SAS is the common system of choice.

Main Modules: We envision the User Interface as having 6

modules: CATALOGS, Doc, DATA, UTILITIES, REPORTS, and HELP. A

breakdown of their functions is as follows:

CATALOGS - view all catalogs

search for data by keyword, category, project, principal investigator

report/print from catalog

create lists of study codes for input into programs

DOC display titles, formats and current FSDB datasets for a given study

view, edit, add, (delete), print abstracts, formats,

and codes

DATA archival/retrieval

UTILITIES - interface to SAS and other analysis systems (by automatically writing input programs for data sets

from their format files)

- generate keypunch setups and "datalimit" checks from documentation files

REPORTS provide various "selectable" reports

HELP context-sensitive help screens

Network Access:

At present the Forest Science Data Bank (FSDB) is accessible to IBM compatibles with access to the Oregon State University Novell Network. This includes all the PC's connected to the 40-plus file servers across the OSU campus as well as users dialing in over 2400 baud phone lines. Unfortunately, at the present time, the FSDB User Interface is not accessible via the Internet. We are, though, exploring the possibility of providing a networked PC which could be remotely operated through a telnet connection. Although the User Interface is not internet-accessible,

it is possible to extract files from

25

our network via the Internet with ftp (file transfer protocol). The Internet address of our server is 128.193.113.1.

Current status

Data: We currently have over 1800 datasets and about 200 NB of data. Roughly half of the data (i.e. 92 MB of data) pertain to LTER. The breakdown by data category is given in Table 1. Every data category contains both current LTER datasets as well as older, pre-LTER data.

Table 1: Shares of LTER data in the Forest Science Data Bank (in % of total 200 MB)

Data Category	%
Streamflow	18
Terrestrial	12
Meteorology	8
Water Chemistry	2
Geomorphology	3
Miscellaneous	2
Other Aquatic	1

All data have been moved from mainframe tapes to the FSDB LAN-server. We are now systematically reviewing all LTER core data sets for sound structure and documentation before making them available to users. The review of meteorological, hydrological, and water-chemistry data will be completed by fall 1989. Reference Stand data will become available in a completely revised form by the end of this year.

User Interface: The catalog section of the User Interface is nearing completion. Work has started on some utilities of the DATA and DOC modules. We have no exact completion date but anticipate that this will be a major focus of LTER III. Clearly, release time for programming support needs to be recognized early in LTER III and planned for accordingly. As modules and individual utilities become completed, we will make them available.

Literature Cited

- National Science Foundation. 1984. Data Management at Biological Field Stations. Report of a Workshop, May 17-20, 1982. W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI. 43 pp.
- Stafford, S. G., P. B. Alaback, K. L. Waddell, R. L. Slagle. 1986. Data management procedures in ecological research. In: Michener, William K., ed. Research data management in the ecological sciences: Proceedings of the symposium; 1984 November 4 -6; Georgetown, SC. Columbia, SC: University of South Carolina Press; 73-91.
- Stafford, S. G., G. Spycher, M.W. Klopsch. 1988 Evolution of the forest science data bank. *Journal of Forestry*. 86(9): 50-51.

Appendix: FSDB catalog and Documentation File Structures

CATALOGS

Projects - List of Research Projects

PROCEDURE C 7 key

PRONAME C 70

FUNDING1 C 15

FUNDING2 C 15

Categories - List of Study Categories

CATCODE C 2 key

NRSTUDIES N 4

NEXTSTUDY C 8 next available study code

CATNAME C 69

PI - List of principal investigators

LNAME C 12 key

FINAME C 10

MINAME C 2

TITLE C 50

INSTITU C 30

DEPARTMENT C 45

PROFAREA C 40 professional area(s) of specialization

TELAREA C 4

TELENUMBER1 C 10

TELENUMBER2 C 10

Keywords - List of keywords

NST N 3 number of studies

KEYWORD C 24 key

STCODES C 240 long string containing studycodes associated

with keyword

Catalogs cont.

Studies - List of studies.

STCODE C 8 key

TITLE1 C 70

TITLE2 C 70

CATCODE C 2 key (category)

PROCODE C 6 key (project)

PICODE C 12 key (pi)

ONLINE C 1 online indicator

LTERCORE C 1 lter 'core' indicator (temporary)

DATABEGIN C 8 start of data collection

DATAEND C 8 end of data collection

OTHERRES C 60 other researchers

STATUS C 1 status code

STORAGE C 8 storage locations

ABSIND C 1 temporary indicators during transition

FMTIND C 1

CODEIND C 1

FILEIND C 1

Formats - List of all formats

STCODE C 8 key

FORMAT C 7 key

OLDCODE C 10 reference to older studycodes

DSNAME C 14 filename

TITLE C 60 format title

CONTACT C 15 contact person

Datafiles - List of all data files

STCODE C 8 key

FORMAT C 3 key

FILENAME C 12 key

FILETITLE C 50

LASTARCH D 8 date of last archival

SIZE N 10

SOURCE C I

DOCUMENTATION

Format - Filenames are 'Studycode' + 'Format number' + '.fmt'

EXECUTIVE SUMMARY

STCODE C 8
FORMAT C 3
VARNAME C 8
VARIABLE C 30
VARDEF1 C 76
VARDEF2 C 76
BLANKS C 2
VARTYPE C 1
VARLEN C 2
DECIMALS C 2
UNITS C 15
MISSING C 8
PRECISION C 15
CODED C 1

Code - Filenames are 'Studycode' + '.cod'

STCODE C 8
VARNAME C 8
CODE C 8 CODEDEF1 C 76
CODEDEF2 C 76

Abstract - Filenames are 'Studycode' + '.abs'

We have opted to use ascii files (max. 75 characters per line) to store abstracts covering the usual abstract topics (study purpose, locations, study design, experimental methods, etc). Abstracts viewed on screen or written to printable ascii files will include additional information from the catalogs.