

01001100 01010100 01000101 01010010

L T E R D A T A B I T S

01001100 01010100 01000101 01010010

Summer 1993, Scientific Information Management Newsletter of Long-term Ecological Research

Welcome to the summer edition of DATABITS. There have been some exciting developments in the use of Interactive Data Access Systems within the LTER network since the last DATABITS issue. This issue includes feature articles on the new All-Site Bibliography, GOPHER information server software, connecting remote PC's to the Internet and a look at data management at Luquillo LTER and site flashes. There have been some changes on DATABITS availability. Among the most exciting is that on-line searching capabilities are now available and all the past issues are now available in ASCII and Postscript forms over the Internet.

CONTENTS

Site Flashes	1
Short News Items	3
LTER Internet Connectivity Corner - PC-NFS SLIP revisited	3
Putting a Gopher to Work	4
LTERnet All-Site Bibliography Project- Phase 1	5
A Structure to Data Management at Luquillo LTER	8
Publication Information	9
1993 U.S. LTER Data Management Workshop Agenda	10

Site Flashes

AND -- The Andrews LTER group has been exceptionally busy in recent months. Susan Stafford (AND) was involved with the planning and organization of the Environmental Information Management Symposium in Albuquerque along with James Brunt (SEV), and Bill Michener (J.W.Jones Ecological Research Center). The Andrews LTER was well represented with Susan, Warren Cohen, and Gay Bradshaw all presenting papers, and Gody Spycher and Don Henshaw also attending.

The Andrews also had their NSF review in June. Tom Callahan (NSF) led a team of eight reviewers that included Bill Michener. The Quantitative Sciences Group and the Forest Science Data Bank was commended in the area of data management in the closing remarks of the review team. Mark Harmon's long term research and renewed modelling efforts at the Andrews LTER were also noted accomplishments.

As if this spring was not busy enough, LTER investigators Fred Swanson, Gordon Grant, and Tom Spies spent most of

two months helping hammer out a Timber/Old-growth/Spotted-Owl/Salmon Forest Management Plan for review by President Clinton and other Cabinet members. Also of note was the opening of the 5000 square foot laboratory at the Andrews field site. The junker trailer days are gone forever. -- Don Henshaw, Andrews Forest LTER, DHenshaw@lter-net.edu

KBS -- KBS GIS is acquiring new aerial photos for the KBS properties. Abrams Aerial Survey of Lansing, Mi. was contracted to do the flyover and prepare B/W contact prints of the site. I have received and reviewed the prints from Abrams and they look good. I have ordered 36x36 enlargements and mylars at scale 1:5500 which are to be prepared from master negatives containing predefined reference points. I chose a minimum of 12 reference points for each print basing my decision on the fact that a minimum of eight points with known coordinates must be identified in order to register the aerials into our GIS. By having Abrams prepare a master negative, I am assured that everyone receiving an enlargement copy is able to register the aerial in a standard manner. I will also be making arrangements with Abrams for a midsummer flight using IR film. I will be using the B/W aerials to update our area basemaps while the IR's will serve to supplement our vegetation database. Arc/Info software will be used for digitizing and mapping. Version 6.1.1 for the VAX has been recently delivered and I am planning to do the upgrade from version 5.0.2 in the next few days.

KBS is also currently operating two weather stations. Campbell CR10 dataloggers are used to record measures such as precipitation, soil and air temperatures, relative humidity, wind speed/direction and solar radiation. Procedures are in place which allow daily means from both stations to be automatically distributed via email to interested parties. Comparing data from both stations allow us to

more readily screen the values for possible instrument malfunction. I am reviewing the climate data from both stations to determine the acceptable level of variation between and within the two sites. I will use the results to define a series of flags which will allow the user to assess data creditability and fitness of use.

Other ongoing data management activities at KBS include the design and development of an on-line KBS LTER Data Yearbook. The concept behind the yearbook is to allow current and future investigators to readily retrieve data, descriptors, summary statistics and/or reference information on any of the core sampling variates in a timely and efficient manner. I have prepared a hardcopy working draft of summary statistics and graphs for 17 of these variates. Summary statistics and graphs provide a measure of data quality control. By evaluating procedures used to create the working draft, I hope to establish links between data files and descriptors which when called will define the parameters necessary for automated standardized statistical output. Automation will in turn provide for consistency and continuity in data format and representation between and within labs. -- *Lolita Krievs, Kellogg Biological Station LTER, LKrievs@lternet.edu*

LTERnet-- LTERnet now has a direct link via fiber optic cable (FDDI) to the NSFnet T3-backbone, and has been upgraded from a Sparc 2 to a Sparc 10 workstation. The 100MBit/sec FDDI link has a separate IP number (192.42.145.13) and domain name (LTERnet-T3), which resides next to the venerable ethernet interface (LTERnet, 10MBit/sec) on the Sparc 10 workstation. LTERnet, with its functions as a communication and information server, hosts the archive for LTER site satellite images on a 10GByte rewritable optical jukebox. With the need to transfer image files of several hundred Megabytes, the 10MBit/sec ethernet interface presented severe limits when compared with the 45MBit/sec speed of the NSFnet T3 backbone.

The other main computer of the LTER Network Support System, Space, has also been upgraded to a Sparc 10 machine. These upgrades have removed bottlenecks in CPU speed and data transfer speed. You can try LTERnet's high-speed interface by connecting to LTERnet-T3.LTERnet.-Washington.edu. Please get in touch with me if you have any questions.-- *Rudolf Nottrott, LTER Network Office, RNot-
t@LTERnet.edu*

NTL -- We have recently purchased Oracle database software for our LAN. The present configuration has Oracle Server on the file server running Novell Netware and the client software, Oracle Client Tools, on the data management staff's Macintoshes. We have the software running and the TCP/IP connection in place and are ready to start creating tables and loading data.

We went through a period of researching the alternative database software available. Important considerations were our heterogeneous group of end users (Macintosh, PC), the

availability of end user software to make retrieving data from the database easy, and software support costs.

Another exciting addition to our LAN was a Polaroid CI/5000 slide maker. It is accessible from both the Macintosh and PC environments. Our new network administrator, Paul Hanson, ran a training workshop for people in the lab. Everyone has been very satisfied with the quality of the slides produced.

We are gearing up for our site visit on July 12-15. John Briggs is a member of the review team and we are looking forward to discussing research data management issues with him. -- *Barbara Benson, North Temperate Lakes LTER, BBenson@lternet.edu*

NWT-- We have made a number of important software acquisitions for our Sun SPARC 2 during the past 6 months: FrameMaker, Asterx, and INGRES/ABF. Hardware acquisitions that will upgrade our data management and modeling laboratory include an additional 16-MB memory and 2.3-GB storage for our Sun SPARC 2, as well as 4 X-terminals. Our GIS laboratory has recently acquired SPlus statistical software and a 1-GB optical disk drive.

We have made a great deal of progress in terms of our transition from an investigator-oriented to a centralized data management system. Two students (full-time in the summer and half-time during the academic year) assist our data manager in the manual entry and processing of data. We meet with investigators and graduate students to design their field and laboratory data collection forms, write Easy-Entry programs for data entry, and write shell scripts for additional processing and archiving of the data. All of our procedures and programs are fully documented in two 3-ring binders that are constantly being updated.

Metadata for all NWT LTER data sets are now stored in an INGRES database. Moreover, we now have nearly 40 data sets archived in ASCII format with the metadata. We have used a modified version of the format described by Conley and Brunt (1991) in *Coenoses* 6. -- *Rick Ingersoll, Niwot Ridge LTER, RIngersoll@lternet.edu*

PAL-- The second field season from September92 to March93 was completed with continuation of the oceanographic, krill and bird surveys. In addition to the weekly on-station sampling, there were two major ship cruises during which core measurements were made off the Antarctic Peninsula. The Iterjan93 cruise occupied stations on five transects. The Itermar93 cruise is a fall cruise which is planned once every six year granting period in order to cover the entire Itergrid of ten transects providing a large scale hydrographic survey of the region in addition to the other core measurements. Preparations are underway for a Spring cruise in August93.

The data structure underwent further organization and standardization in order to promote more rapid documentation and data report generation. Progress was made on

definition of dataforms and core data sets. The on-line browse program work continued in parallel with implementation of a gopher site organization. A new 500mbyte disk came on-line to replace old hardware. A remote campus site appletalk network was linked to the central lter appletalk network using cayman systems gatorboxes performing tunneling across the campus internet. This permits both apple file sharing of remote mac/central mac and remote mac/ central unix disks as well as the ability to print remotely on postscript and mac printers. In addition, remote site as well as local mac backups are being done routinely onto an 8mm tape drive from one central location making use of the gatorbox to gatorbox tunneling.

Preparations have begun in order to facilitate dataset documentation while at sea. Also, two unix workstations are being prepared for at sea use. This will eventually permit the palmer lter database of information to be available in the field. -- *Karen Baker, Palmer LTER, KBaker@lternet.edu*

VCR -- We have been working hard to address the information needs of our wide-ranging PI's. Currently we have active investigators at more than 6 institutions, so keeping everyone informed of VCR activities was getting to be a real problem. Our solution has been to institute an electronic calendar system and make site documents available on-line. The calendar system, which sends daily and weekly notices of LTER activities to all VCR PI's and students via electronic mail, was discussed in the last issue of DATABITS and has been a smashing success. It has also been and a major contributing factor to the 100% electronic mail connectivity within the site as several investigators obtained electronic mail specifically for calendar access. Our new on-line information system, the Virginia Coast Reserve Information System (VCRIS) has been a similar success. It gives PI's at remote institutions the same access to documents and data as PIs at the University of Virginia, where the VCRIS is located. VCRIS uses the GOPHER software and is discussed extensively elsewhere in this issue of DATABITS.

NASA has installed a sun photometer at the site for use in the ongoing SCAR-A experiment. Data from LTER automated meteorological stations will be used to maximize utility of sun photometer data. We have also been working at making an increasing amount of our field data (especially those from automated monitoring devices) come in an electronic form and are experimenting with using bar-codes for collecting some types of data. We have been experimenting with integrated sensor/logger units which measure a single environmental parameter, but are small (and cheap!) enough to deploy in widely separated locations.-- *John Porter, Virginia Coast Reserve LTER, JPorter@lternet.edu*

Short News Items

Environmental Information Management Symposium held

The University of New Mexico hosted a symposium on Environmental Information Management May 19-22, 1993. The symposium was supported by a grant from the National Science Foundation to symposium organizers (and LTER Data Managers) James Brunt, William Michener, and Susan Stafford. The symposium included papers by a wide array of information professionals and a published copy of the symposium papers is expected out early next year. Researchers from LTER sites were much in evidence, with sessions chaired almost exclusively by data managers from LTER sites.

National Center for Synthesis in Environmental Biology

Work is moving forward on the establishment of a National Center for Synthesis in Environmental Biology. A panel, chaired by Dan Sulzbach of the San Diego Supercomputing Center, is completing a report on the facilities, both technical and physical, needed to support such a center. The proposed center will need facilities to support several types of activities, from intensive workshops and conferences, to individual research. As the name implies, the center will focus its efforts on the synthesis of existing data, rather than the collection of new data. The LTER program was well represented on the panel. Participants included John Vande Castle and Rudolf Nottrott from the LTER Network Office and James Brunt and John Porter from the SEV and VCR LTER sites.

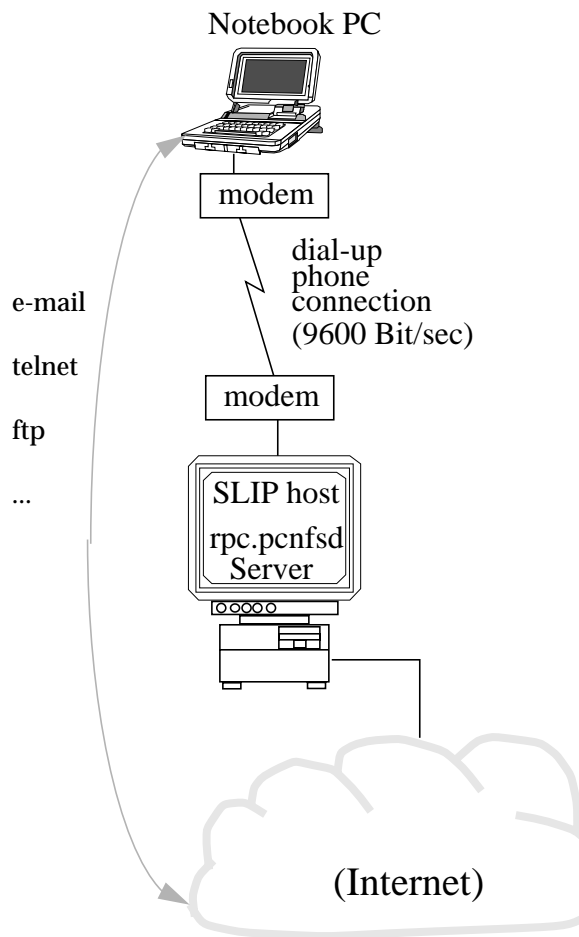
GPS Workshop Postponed

The LTER GPS workshop, originally planned for August 1993 has been tentatively rescheduled following the All-Scientists' Meeting. Several factors, notably schedule conflicts for the instructors, technical problems in administering funds and higher than expected costs in upgrading the LTER GPS units led to the rescheduling. New advanced GPS processing software is now available for use with the LTER GPS units.

LTER Internet Connectivity Corner - PC-NFS SLIP revisited

In a past issue of DataBits, I reported on our efforts to use PC-NFS, from Sun Microsystems, to integrate DOS-based PCs with Unix-based workstations. We have now been through several versions of PC-NFS and have found that it can provide transparent network printing and file sharing for PC's in an ethernet LAN TCP/IP environment.

Figure 1: Dial-up SLIP connection



This scheme ends, however, when you are traveling with your notebook PC, and when the only link to the world of networks can come over the hotel phone. While it is easy to use “plain” communications software, such as Procomm, to turn your PC into a terminal attached to a remote computer, you need more if you want functions such as e-mail, FTP and Gopher to put messages or files directly onto your PC - you need SLIP or PPP. For example, if you read your e-mail by using your PC as a terminal on a remote computer, then all your messages stay on that remote computer. You can only read them when you are hooked up to the phone. In contrast, if your PC is a node on the Internet, the messages will be transferred to your PC and you can read them anytime you like, even when you are off-line (off the phone). The prospect that I could take a PC with my whole backlog of e-mail messages on the next camping trip made me look at Serial Line IP (SLIP) again.

PC-NFS has a SLIP network driver, which replaces the ethernet driver when I am away from the office (Figure 1). I had much trouble getting the SLIP drivers to work initially, and still not everything is working perfectly under Microsoft Windows, but the three Internet applications I wanted most are now on my PC even when I only have a phone. The applications are FTP, Telnet and e-mail (I use Eudora). This article went to the editor in Virginia directly from my PC, using FTP over the phone. And I am now writing my e-mail

messages on my PC, then later I briefly call the SLIP server to connect to the Internet and send the messages, as well as receive new messages.

Sun has promised an implementation of Point-to-Point Protocol (PPP, a successor to SLIP), in their next release of PC-NFS. I am also looking at more mature SLIP products from other sources (FTP Software and Netmanage, and public domain software). SLIP and PPP are slower than a “plain” modem connection due to the overhead in their network layer packets. Therefore, you need modem speeds of at least 9600Bits/sec and these modems are just becoming affordable. Nevertheless, if trends continue at the present rate, many of you may have a serial Internet protocol stack on your portable PC in the near future.

I am interested in hearing from any of you about related efforts at gaining TCP/IP connectivity “at the slow end”. --
Rudolf Nottrott, LTER Network Office, RNott@LTERnet.edu

Putting a Gopher to Work

GOPHER: 1) a small rodent who lives in a hole in the ground. 2) a computer program to allow easy, menu-driven access to information.

What's the marine weather forecast for the Virginia Coast? What was the weather like on October 29, 1991 on Hog Island? What publications of Virginia Coast Reserve LTER researchers had the word “island” in the title? Finally, what did president Clinton say at his last press conference? What do these questions have in common? Easy, they can all be answered by using VCRIS, the Virginia Coast Reserve Information Server which is based on the GOPHER software from the University of Minnesota (home of the Golden Gophers).

First, lets talk a little about how GOPHER operates at the level of the user, because one of its real strengths is that GOPHER is so simple to operate that new users can teach themselves. To start a gopher client, a user simply runs the program (usually named “gopher”) from their computer. GOPHER client's exist for a wide number of platforms (UNIX, PC, MacIntosh, VMS, and IBM-CMS) and what you then see depends on the platform you are using. The only limitation is that, to get the full power of GOPHER, the client must run on a computer connected to the Internet. A common way to do this is to dial-in to a UNIX or other multiuser computer and utilize its networking capabilities. On a terminal hooked to a UNIX computer, you will see a menu appear with a pointer that can be moved using cursor keys (Figure 2). On a MacIntosh running “TurboGopher” you will see (in addition to a tiny dancing gopher) the same menu, but mouse-selectable. If you select an item on the menu by hitting “enter” (or clicking a mouse) you will either have a text file displayed on your screen or a new sub-menu, which may itself have text files or new menus. Mov-

ing in the menu system or quitting is easy -- hitting 'u' moves you up a level in the menu system and hitting 'q' quits GOPHER. Using GOPHER is easy, because there are basically just 4 commands: UP, DOWN, GOTO and QUIT!

Now lets take a more sophisticated look at a GOPHER menu. One of the major virtues of GOPHER is that it hides real power behind simplicity! When you run a GOPHER client, it connects to a GOPHER server, running on a UNIX or MacIntosh computer. The particular server selected is specified in the configuration section of the software or inserted at compilation time so the typical user never sees it. The client constructs menus based on the information provided by the server. That information can be in two forms: data that actually reside on that particular server and links to data on other servers (both GOPHER and otherwise). A user at the client level can't distinguish which is which! If a link is activated by selecting the associated menu item, a connection between the client and the server for that particular item is instigated.-- *John Porter, Virginia Coast Reserve LTER*

Figure 2: Sample GOPHER menu

```
Internet Gopher Information Client v1.11
Root gopher server: atlantic.evsc.Virginia.EDU

--> 1. About the Virginia Coast Reserve Information System/
    2. Calendars/
    3. Data/
    4. Opportunities/
    5. Other Gophers and Net Information/
    6. Personnel/
    7. Publications/
    8. Weather: Central Virginia .
    9. Weather: Coastal Marine Fest (KWBC).
   10. Weather: Hampton Roads .
   11. Weather: Radar Natl Ascii Maps.
   12. Weather: Virginia /

Press ? for Help, q to Quit, u to go up a menu      Page: 1/1
```

LTERnet All-Site Bibliography Project- Phase 1 INTRODUCTION

The goal of this project is to allow a person at any LTER site to search the LTER-related bibliographic records from all sites. Possible queries might be

- What has been published on ground squirrels?
- What has Dr. Joan Martin done on prairie grasses since 1987?

I carried out phase 1 of this project under the guidance of Caroline Bledsoe and with the collaboration of Rudolf Notrott.

There is not yet a completely satisfactory solution to the project's goal. Fortunately, however, during the last few years, some remarkable work has been progressing that makes it easy to provide a limited solution. These are "WAIS" and "gopher" - public domain, information access programs that work over a computer network. Furthermore, the academic biology community has embraced

them, with a growing number of on-line information sources covering, for example, biological diversity, taxonomy, ecology, and currently active grants.

Phase 1 implemented a centralized, all-site bibliography using gopher and WAIS, located at the LTER network office in Washington state. Some details are described below.

GOPHER AND WAIS

WAIS-- WAIS (pronounced "ways") stands for Wide-Area Information Servers. Its intent is to permit the user to search in an easy and consistent manner for relevant information from a variety of sources stored on computers. WAIS, as well as gopher, uses the "client-server" model of operation. A WAIS client allows a user to ask "questions" about "documents" in a database (a WAIS "source"). A WAIS server provides answers to these questions. The answer to a question is a list of documents that match the question, ranked by how well they match. The user may ask to view any document in the list.

Any word in a text document is searchable with WAIS. Therefore, there is no concern over whether a limited, prescribed set of keywords is sufficiently inclusive to return all appropriate references. However, this depends, too, on the nature of the document. The title of an article may not be explicit enough to catch the full set of subject areas covered by the article; abstracts in the bibliographies would alleviate much of this problem.

With a simple WAIS client, a question consists of a set of search words and the source (database) to be searched. This is the form of searching used from gopher, see below.) There is a more sophisticated type of WAIS client, but it does not presently work with gopher.

Gopher-- Gopher is an interactive, network navigation tool. Its purpose is to hide the networking details of getting to other computers and searching for and retrieving publicly-accessible resources located there. It also uses the client-server model of operation. The gopher clients use a simple, menu-driven interface.

Gopher is designed to work hand-in-glove with WAIS when the appropriate activity is to search documents based on a set of search words. This hand-off of the search to WAIS is transparent to the user of gopher.

Gopher and WAIS clients are currently available for UNIX, the Mac, DOS PCs, and probably VMS.

CONSTRUCTING THE ALL-SITE BIBLIOGRAPHY

Each bibliographic entry, or record, comprises a number of types of information, such as author(s), title, journal name, volume number, and abstract. This information is more useful, at least in a computer sense, if it is separated into distinct fields. The format for the LTER on-line bibliography explicitly labels each field in the record.

The site bibliographies are maintained using a variety of software, ranging from word-processors to full bibliographic database systems. Most database programs and word-processors have some way to save (or “dump”) the information from its binary database or a word-processing file as plain text.

My task in phase 1 was to homogenize, as much as possible, the various formats of the text dumps into a single display format that is consistent and useful. I call this “the LTER on-line format”. I wrote a set of programs that filter these text files. Essentially what the filters do is to rearrange the text of the input into the common format, also adding field labels. The filters are written in the “awk” programming language because it is designed to do exactly this kind of text manipulation. Ideally, the on-line format has each type of information divided into separate fields, each on a separate line.

The LTER On-line Format

I followed three guidelines in developing the on-line format.

(1) When a person reads this format, there should be no mistaking what each field contains. (Yes, Murphy guarantees no matter how clearly everything is labeled, someone will misunderstand.) Success here is a balance between clarity and conciseness.

(2) Gopher allows the retrieved information (bibliographic records in this case) to be saved to the user’s computer. Should the user want to import these saved records into their own bibliographic database system, a consistent, labeled format makes this easier.

(3) When gopher and WAIS are capable of sophisticated, database-style searches, this labeled format will more easily permit these improved searches on the LTER bibliography.

Here is a sample record in this format:

```
author :: Bisson, Peter A.; Quinn, Thomas P.; Reeves,
         Gordon H.; Gregory, Stanley V.
title :: Best management practices, cumulative effects,
         and long-term trends in fish abundance in
         Pacific Northwest river systems
date :: 1992
pages :: 189-232
phrase :: In
editor :: Naiman, Robert J.
book title :: Watershed management - balancing
            sustainability and
            environmental change
publisher :: Springer-Verlag
place of pub. :: New York
key words :: LTER92; HJA3; RWU-4356; FY92
document type :: book
LTER site :: AND - H.J. Andrews Experimental Forest, Oregon
LTER bdbs :: Pro-Cite
LTER bdbs rec :: 1357
```

“bdbs” abbreviates bibliographic database system. “bdbs rec” is the identifying record number used in the bibliographic system.

Construction Difficulties

It is easy to extract each field from a database system, such as from Pro-Cite for the example above. Pro-Cite has up to 45 fields available for partitioning the information in a bibliographic record. On the other hand, not all systems or their users divide the information in the same way or to the same degree.

In some of the text versions it is quite difficult to extract each type of information, because all the information is lumped into a single field. As can be seen in the following example, there is no sure, easy way to determine programmatically where one field ends and the next begins, especially between the title and the name of the journal:

7.Allison, T.D. 1992. The influence of deer browsing on the reproductive biology of Canada yew (*Taxus canadensis* Marsh.) III. Sex expression. *Oecologia* 89:223-228.

The structures of some bibliographies are intermediate; their information is partially lumped. An example comes from another of the site bibliographies:

(1). Author(s): Collins, S.L., and D.J. Gibson

(2). Year: (1990)

(3). Title: Effects of fire on plant community structure in tallgrass prairie.

(4). Editor(s) and Publisher: In S.L. Collins and L.L. Wallace (eds): Effects of Fire on Tallgrass Prairie Ecosystems. Norman, OK: University of Oklahoma Press,

(5). Volume and Page(s): pp. 81-98.

A hard problem I did not attempt to solve is dealing with mistakes in the records. These include information in the wrong field, such as the date in the author field, and typographic errors.

Problems with WAIS

The WAIS Indexer-- WAIS has an indexing program which creates the actual form of the database used by the server. A problem with the current WAIS indexer is dealing properly with the format of one’s own particular documents, in this case, bibliographic records.

After searching the bibliography, gopher returns a list of matching records, which is composed of “headlines” - much like the headlines in a newspaper. Each headline is constructed from its corresponding document. Presently, the headline is just the first line of the record. For the on-line format shown above, this would be the authors field, not the most helpful possibility.

For better headlines, the filter programs add a new first line composed of the primary author’s last name and as much of

the title as will fit on one line. Unfortunately, this new line clutters the record with redundant information (and it usually ends in mid-word). To improve readability, a second line, composed of just the separator (“::”) is added after the headline.

So, the Bisson, “et al.”, record now begins like this:

```
ID :: Bisson: Best management practices, cumulative effects,
::
author :: Bisson, Peter A.; Quinn, Thomas P.; Reeves,
        Gordon H.; Gregory, Stanley V.
title :: Best management practices, cumulative effects,
        and long-term trends in fish abundance in
        Pacific Northwest river systems
```

Gopher and WAIS are still under development. I anticipate easier customization of the indexing program will permit doing away with the “ID” line of each record.

WAIS Searches-- WAIS does not currently support the sophisticated type of search available from most commercial, bibliographic database software (e.g., Pro-Cite), such as “find all works by author X on subject Y published after date Z”.

In fact, WAIS does not even allow simple Boolean combinations of search words. For example, one would like to use the search phrase “X AND Y” to come as close as possible to asking for works “by author X” (“and”) “on subject Y”. Presently, the phrase “X AND Y” will return documents containing only the word “X” and those containing only “Y” and those containing both “X” and “Y” (and perhaps all those containing the word “and”). Documents with both “X” and “Y” will be ranked higher than ones with just one of the two, so not all hope is lost.

A Context for Phase 1

As the technical and legal kinks are worked out of making computerized information easier to obtain, the usefulness of the on-line LTER system should increase dramatically. This includes first adding abstracts and then the entire text of each article available, along with any figures and photographs (a PostScript version of each article is a likely adjunct).

Some data sets are presently on-line at LTERnet. Not-so-distant future enhancements to gopher and WAIS will allow searches that start out with a question like “Show me what has been published on this topic”. After reading a number of the abstracts, you would instruct the system, “Show me all the articles in the system that are similar to records 4 and 27”. After looking them over, you might then decide to print the articles for records 4 and 162, complete with tables and figures, to your local printer and have the corresponding raw data for the two articles transferred to your computer.

Acronyms

- acronym Annoying, Compressed Representations Of Names You (always) Misunderstand
- awk A. V. Aho, P. J. Weinberger, B. W. Kernighan
- DOS Devil’s Operating System
- gopher (not an acronym)
- IBM I’ve Been Moved
- LTER Long-Term Ecological Research
- NCSA The National Center for Super-computing Applications
- PC Personal Computer (when applied to a class of desktop machines)
- PC Personal Confuser (when applied to DOS machines)
- UCD University of California, Davis (The Farm)
- UNIX (not an acronym; a real operating system)
- veronica Very Easy Rodent-Oriented Net-wide Index to Computerized Archives

WAIS Wide Area Information Servers

Existing WAIS and Gopher Servers.-- Thinking Machines, the main sponsor for the creation of WAIS, has a list of nearly 400 publicly advertised WAIS sources, covering a diverse array of topics, such as

- LTER Information System
- VCR Information System
- weather maps and reports
- the NOAA National Environmental Data Referral Service (NEDRES)
- DOE climate data
- the NSF publications database
- an archive of postings to various Usenet and Bionet/ BIOSCI electronic discussion groups related to environmental issues

There are also many gopher servers. Here is a partial list of those of interest to biologists:

- sparc.ecology.uga.edu CWT - Coweeta Hydrologic Laboratory
- merlot.welch.jhu.edu Gopher-to-ftp access to “The-Scientist”. Gopher service to help find other scientists and projects
- ftp.bio.indiana.edu The IUBio archive. Genbank(tm), a lot of biological software, WAIS code modified to support Boolean searches
- huh.harvard.edu The Biodiversity and Biological Collections Gopher
- gopher.ncc.go.jp National Cancer Center (Tokyo Japan) for biological research work and cancer treatment
- ecosys.drdr.virginia.edu an educational EcoGopher
- menu.crc.ac.uk Bio-Gopher to serve the medical research community in general and the genomic researchers in particular
- ftp.bchs.uh.edu The UH PIR (protein sequence data) Gopher hole
- bioftp.unibas.ch Prosite, EMBL data, and lots more
- gopher.micro.umn.edu The home gopher hole

There is also a gopher information system named Veronica, the purpose of which is to help one locate for gopher servers that contain resources of interest.

- futiq ue.scs.unr.edu Veronica offers a keyword search of most gopher-server menu titles in the entire gopher web -- Harvey Chinn, Graduate Group in Ecology, University of California, Davis

A Structure to Data Management at Luquillo LTER

Ever since 1989, the LTER Data Management (DM) at the Luquillo Experimental Forest, has been involved in the development of its structure. Our motivation has always been a clear consciousness of the importance of the gathering and safe archival of ecological data. Long-term records of the data serve our present researchers in their endeavor to learn more about our planet, and will serve future scientists to continue our commitment to endure and conserve Her. The main responsibility of Data Management is to make sure that the information is carefully protected and made available to the scientific community. In order to achieve this important task, the structure developed pays close attention to the every day tasks of entering, checking, distributing and archiving data in an organized manner. In the past four years, DM have gone through the process of learning, along with the LUQ scientific community, what are the best methods and procedures to obtain the best results in analyzing their data. Figure 3 shows our perception, concepts and structure of Data Management.

Figure 3: Data Management Functions

BASIC			
ENTRY	DOCUMENTATION	ARCHIVAL	RETRIEVAL & DISTRIBUTION
y ENTER	y HELP DOCUMENT	y BACKUP	y ANSWER TO
y CHECK	y CATALOG	y PROTECT	y REQUESTS
y REPORT	y FOLLOW UP	y ARCHIVE	y REPORT
INTERMEDIATE			
MANIPULATION	ACQUISITION		
. REFORMAT	y OBTAIN		
y REDUCE	y REFORMAT		
	y CHECK		
	y DISTRIBUTE		
ADVANCED			
ANALYSIS	PUBLICATION	PRESENTATION	

Basic Functions

DATA ENTRY. This is the function that makes data available to the user. The idea is to facilitate data manipulation processes and to assist in the storage and future retrieval of data. This function involves entering the data ON on magnetic media, using quality control procedures (QC) to check the data entered, and producing reports to the investigator

in charge of the data for further QC and manipulation processes.

DATA DOCUMENTATION. Special consideration has been given to data documentation. In each of our annual reports since 1990 we have been including copies of what we called first "Data Management System" (1990, 1991), and later "Data Management System guidelines" (1992). For this year we included a revised version in the annual report called "GUIDELINES: HANDBOOK TO THE DATA MANAGEMENT POLICIES, DATA ARCHIVAL AND REQUESTS". This handbook contains all the protocols, guidelines and forms that have been developed at DM in collaboration with our principal investigators and the Data Management Committee to facilitate the filing of data and requests at DM. The 1993 version also includes a copy of the memo that must be sent to one of our principal investigator when filing a copy of a presentation or publication. We encourage investigators to meet with us to jointly fill out the documentation forms. The information on the documentation forms is particularly useful to DM when we prepare the Catalog of LTER Data Sets. This catalog has been indexed by the LUQ Research area, as they appear in our original proposal for LTER.

DATA ARCHIVAL. We have grouped all the functions of DM that are related to the security and durability of the data sets filed. Regular backups are made after each update of the file. Two copies of the files are saved and stored in different places. The original data sheets and documentation are filed in fire-proofed cabinets to eliminate possible loss in case of catastrophic events. Files with certified data quality will be archived in permanent computer storage medias at the International Institute of Tropical Forestry (IITF) and at the Division of Terrestrial Ecology (TED). The local Data Management Committee met this year to coordinate this activity.

DATA RETRIEVAL & DISTRIBUTION. In the last period (1992 to the present) the total of data requests has reduced but there has been an increase in requests that include further data manipulation after retrieval.

We are also responsible of maintaining our LEF and LTER bibliography and catalogs in computers files for on-line reference and listing. We produce annual reports including updates of our catalogs and lists in order to present the work done at DM and the work done in general by our investigators.

Intermediate Functions

We call these functions "intermediate" because of the level of complexity of the work they involve DM in. They are related to the data manipulation and acquisition of data from other sources. The time, difficulty, and the resources involved in the completion of these tasks depend on the structure of the data set to be manipulated, the analysis the data will be submitted to, and the quality of the data entered.

DATA MANIPULATION. When data needs to be reduced (make summaries and averages, counting) the manipulation of data is usually done using the computer software that was used to enter the data. When data requests involve the change of format of the data entered, additional programs have to be developed to manipulate the data. The data in the files we receive from USGS are useful only after the format of the file is changed. We have developed PARADOX programs to change the format of these data sets. It is very important to observe the significance of good data entry, documentation, and quality control procedures to obtain good results with data manipulation. A good example of this is the manipulation that we have just performed on some of the Landslide Revegetation data sets. The data was first reduced and the resulting files were then re-formatted to a matrix-like file to be processed by a statistical software. The original data sets showed to have good quality in this process and the manipulation was done successfully. This is the result of an early and continuous coordination of the data entry and QC processes between the investigator, L. Walker, and the Data Manager.

DATA ACQUISITION. In our last Scientific Advisory Board Report from the LTER committee it was pointed out that DM "should actively collect information from other sources and present them in a useful compendia". Data Management has been involved in the acquisition of these data since summer, 1992, when we contacted different sources of data gathered at Luquillo. The data obtained come from different sources, namely, USGS, NADP, and NOAA. The variables included are precipitation, max and min temperature, general USGS met station data, streamflow and sediment data. We produce graphs showing these data. As mention before, most of these data files have to be re-formatted in order to be used. The daily streamflow data from eleven stations and the three meteorological stations data have been changed into a usable format.

Future goals

On Figure 3 you can observe that the boxes under "Advanced Functions" do not have bullet items. These are currently done at an individual basis by the investigators (or groups of investigators under each research areas). Data Management collaborates, with data entry, retrieval, and manipulation, to accomplish these functions but does not perform them directly. With proper training, and additional personnel we could get involved in the analysis, publication, and presentation of data. The scientific community of our site has made us aware of their need of support in the analysis of their data for publication and presentation. The statistical background of our DM staff could be classified from basic to intermediate. The load of work and need for further training in statistical analysis, makes it impossible for DM at LUQ to complete such tasks. Nevertheless, we envision future involvement in this area to broaden our development in Data Management. This will depend entirely on our future economical resources to hire more specialized personnel and train the existing ones.

Assisting the investigators in the analysis of their data is not the only way that we can achieve this goal. Even though it is a fact that investigators tend to view data management as a means to more efficient data collection and analysis (Gorentz, 1992), DM at LUQ has developed in the process, the need of communicating the best methods found to manage the data to do analysis. We have in mind James T. Callahan's advice to all Data Managers at the 1992 Data Management Workshop held at Honolulu Hawaii. He emphasized the fact that all Data Managers have publishable material without sometimes being aware of. He encouraged us to get active in the area of research and publishing in data management (Nottrott and Porter, 1992). Data Management at LUQ envisions this as our goal and achievement in the next stage of our existence as an LTER site. We think this activity requires more maturity in the discipline, and probably more resources. We are committed to achieve this goal. -- Eda C. Melendez, Luquillo Experimental Forest LTER

REFERENCES

Waide, R. B, Lugo, A. E. , et. al. 1988. A proposal to the National Science Foundation from the Center for Energy and Environment Research, U.P.R. and the Institute of Tropical Forestry, USDA for the Long-Term Ecological Research on the Luquillo Experimental Forest.

Gorentz, J. B. (ed.). 1992. Chapter 1 in Data Management at Biological Field Stations and Coastal Marine Laboratories: Report of an Invitational Workshop, April, 1990. W.K. Kellogg Biological Station, Michigan State University.

Nottrott, R. and Porter, J. 1992 (eds.). Chapter 2 in Proceedings of the 1992 LTER Data Management Workshop. Long-Term Ecological Research Network Office, Seattle, WA

Publication Information

<i>Editor:</i>	<i>Circulation</i>
<i>John Porter</i>	<i>Rudolf Nottrott</i>
<i>Virginia Coast Reserve LTER</i>	<i>LTER Network Office</i>
<i>Dept. of Environmental Sciences</i>	<i>College of Forest Resources AR-10</i>
<i>Clark Hall, University of Virginia</i>	<i>University of Washington</i>
<i>Charlottesville, VA 22903</i>	<i>Seattle, WA 98195</i>
<i>Phone: (804) 924-8999</i>	<i>(206) 543-8492</i>
<i>Email: jporter@lternet.edu</i>	<i>rnottrott@lternet.edu</i>

DATABITS issues are available as ASCII text on the VCRIS Gopher. Use: gopher atlantic.evsc.virginia.edu to access VCRIS. Postscript copies of DATABITS are available via anonymous FTP to lternet.edu

1993 U.S. LTER Data Management Workshop Agenda

Editors Note: This is a draft agenda and subject to change. The workshop will be held July 29 - 31, 1993, in Madison, WI.

Thursday, 7/29: Arrivals

7:00-9:00 pm Welcome Reception

Introductions and Site Flashes

Friday, 7/30:

7:15 am Breakfast

8:00 am Old Business

- New Sites, guests
- publication arenas for DM-type papers
- ERIM curricula
- All Scientists' Meeting presentation and demos
- hands-on training in data exchange techniques
- common format for data exchange
- relationships with professional societies
- CERN collaboration (Brunt)
- Expansion of LTER Core Dataset Catalog

9:00 am New Business

- NASA Collaboration
- Risser LTER Review Report (Briggs)

10:00-10:15 am Break, Refreshments

- 10:15 - Online Information Services Using Gopher and Wais,
(Porter, Nottrott)
- Information Server on LTERnet, Gopher/WAIS
- All site bibliography project.

- 11:00 - Site Data Management Policies Discussion.
(Porter, Baker?)

12:00-1:00 pm Lunch

Friday Afternoon

- 1:00 - Jim Gosz, Division Director of Environmental Biology
at NSF on data management and networking.

3:00-3:15 pm Break, Refreshments

3:15 - Working Groups Breakout

- 4:00 - Tour and demo at Environmental Remote Sensing Center
at the University of Wisconsin.

7:00-... Dinner (Midwestern Food)

Saturday, 7/31:

7:15 am Breakfast

8:00 am Meeting begins

- 8:15 - Guest Speaker Blanche W. Meeson, Project Manager for
the Pilot Land Data System (PLDS), NASA Goodard Space
Flight Center.

Blanche will give an overview of her work, with special
emphasis on management of spatial data (metadata, exchange standards, integration into
scientific information systems, data access and archiving).

9:00 Discussion

- on subjects presented by Blanche, with a view of finding areas in which
collaboration is desirable and feasible,

- and Management of Spatial Data. What metadata is needed? How should it
best be stored and queried?

10:00-10:15 am Break, Refreshments

10:15 - Working/Discussion Groups - Old and New TBA

12:00-1:00 pm Lunch

1:00 - General Discussion on Working Group Topics

2:00 - Working Group Enhancements / Report Production

3:00-3:15 pm Break, Refreshments

3:15 - Software Demos / Working Groups if Necessary

4:30 pm Wrap up - Action Items Identified

5:00 pm Meeting Adjourns

Sunday, 8/1: ESA Meeting or Departures