CPER LTER

Data Management Report

William K. Lauenroth

Natural Resource Ecology Laboratory
Colorado State University
Fort Collins, Colorado 80523

October 1983

1. BACKGROUND

1.1 Description of the Site and Facilities

Location and Land Management

The Central Plains Experimental Range (CPER) is located in the western division of the Pawnee National Grassland (Fig. 1). The western division of the Pawnee National Grassland is 42,700 ha and the CPER is 6,280 ha.

The CPER is located 19.3 km northeast of Nunn, Colorado, and 40 km south of Cheyenne, Wyoming (Fig. 1). The CPER was established in 1939 to answer questions which were important as a result of the land use of the 1930's. A number of pastures were set aside for long-term experiments. Twelve half-section (129 ha) pastures were assigned four each to heavy, moderate, and light summer grazing. In 1958 two of the replicates were changed to winter grazing. Each of these and several other pastures have at least one exclosure of 0.5 or 2 ha excluding livestock grazing since 1939. Permanent quadrats have been established in these pastures, and in most years composition of vegetation has been measured.

All of the CPER is available for use in the Long-Term Ecological Research Program (LTER), but some is primarily dedicated to studies conducted by the Agricultural Research Service (ARS). The Pawnee National Grassland is available for extensive studies which require a great deal of land area but do not require rigid control. The LTER program staff will assist investigators in securing cooperative agreements with the U.S. Forest Service for use of these lands.

The CPER was set aside as a range research area in 1937 in the northwestern corner of the Weld County Land Utilization Project. The CPER was developed as an experimental area by the U.S. Forest Service in cooperation with the Resettlement Administration. In 1939 the development work was completed and research was started by the Forest Service that spring. Initially, the CPER contained 4,025 ha of federally owned land available for experimental use. In 1940, the Soil Conservation Service and Forest Service signed a memorandum of agreement to conduct research on the CPER. In 1953 the ARS assumed operation of the CPER.

In summary, the following legal relationships exist at the Pawnee National Grassland today. The Pawnee National Grassland is administered as a part of the National Forest System. The CPER is located in the northwestern part of the western portion of the grassland. Of the 6,280 ha of the CPER, 356 ha are privately owned and are made available for use through use-exchange agreements. There are also about 500 ha within the CPER that are privately owned and operated in five separate tracts.

Within the CPER was located the Pawnee Site of the International Biological Program (IBP) Grassland Biome study. In 1968, a cooperative agreement was signed among ARS, CSU, and the IBP's Grassland Biome Program. The agreement permitted IBP to conduct grassland research on a portion of the CPER and provided for mutual cooperation. The agreement also permitted

the construction of needed facilities on the CPER. These included an office-lab-cafeteria, storage shed, dormitory, residence, barn, and corrals. This agreement was amended in 1975 when the IBP program was phased out and is currently the agreement of record.

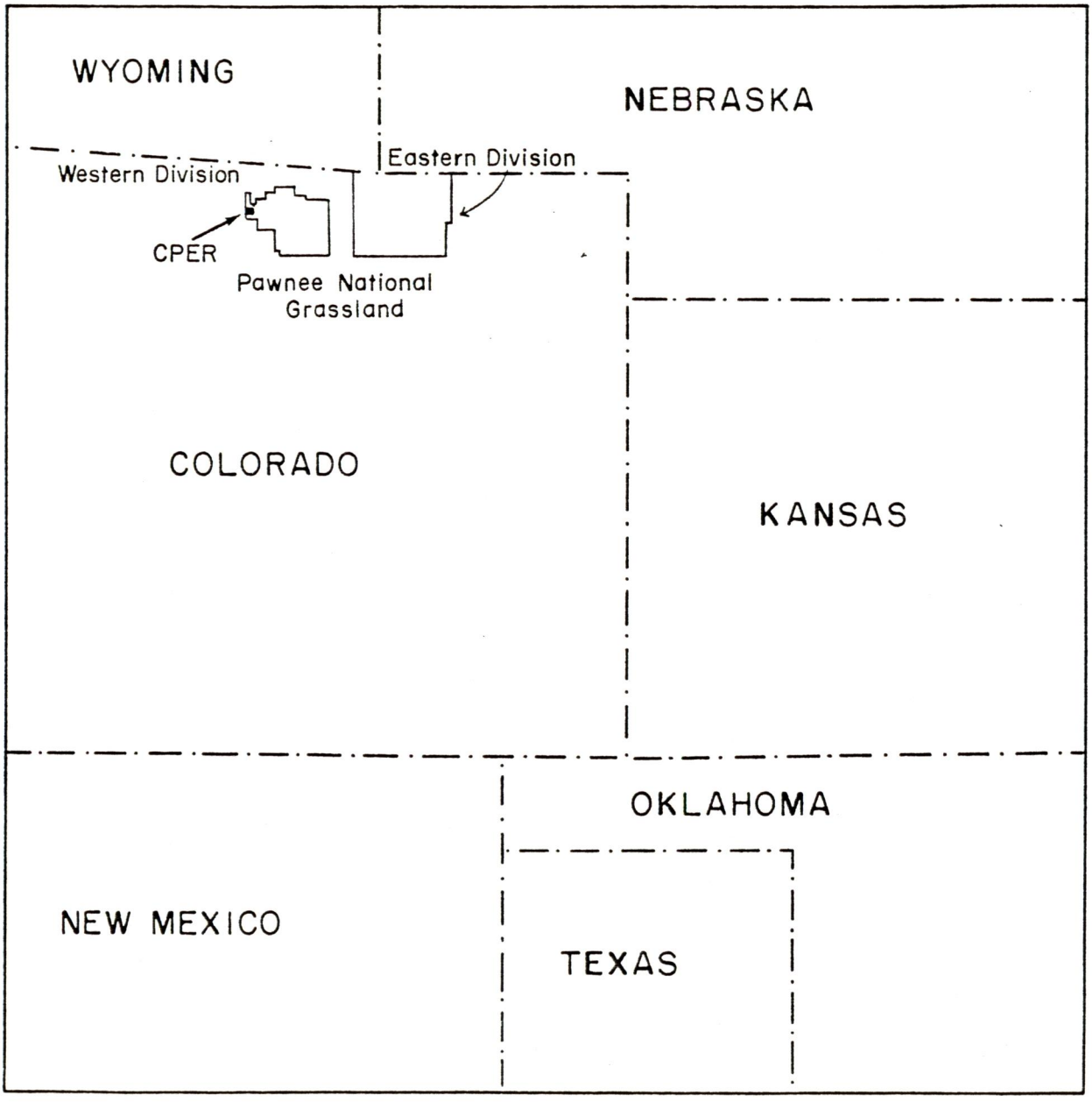


Fig. 1. Map showing the regional location of Pawnee National Grassland.

The scope of research conducted at the CPER Site has changed considerably over the past years. Initial studies on the CPER were concerned primarily with stocking rates, grazing intensities, plant phenology, plant ecology, development of research techniques, time of summer grazing, use of cow manure and commercial fertilizers, effect of time of defoliation of plants and reseeding experiments. In recent years the scope of studies has increased considerably. Studies are concentrated on determining principles of plant phenology, physiology, and ecology. Nutrient needs of range cattle, management practices involving herbicides, fertilizers, growth stimulants, and effective moisture use are other areas of current concern. Experimental areas are varied and lend themselves to many different experimental activities.

1.2 Nature of the Research at the Site

A central concern of the CPER/LTER is the question of the origin and significance of spatial heterogeneity. To address this problem we will be conducting experiments across a range of spatial scales. We have identified a nested hierarchy of units in the steppe landscape ranging from individual slope positions to physiographic units. Responses at each spatial scale will have characteristic ranges of time scales.

The general idea which will guide our LTER research is that the shortgrass steppe is a pulse regulated system. Exogenous and endogenous pulses at a variety of space and time scales provide explanations for the structure and behavior of the system.

Research disciplines involved with the CPER/LTER include geomorphology, soil science, meteorology, hydrology, phytosociology, zoology, entomology, microbiology, range science and systems analysis. At the current time researchers are drawn entirely from the Fort Collins scientific community comprising University, State agency and Federal agency scientists. The CPER/LTER is a joint project of the Range Science Department and the Natural Resource Ecology Laboratory (NREL). The physical campus headquarters of the LTER is located in the Range Science Department. The CPER receives a moderate amount of use by field oriented classes in the College of Forestry and Natural Resources, the College of Agriculture and the College of Natural Sciences.

1.3 Data Bases

Data have been collected at the Central Plains Experimental Range since it was set up in 1939. These data are available primarily in the form of ARS reports; however, weather data are available on tape. Related weather data are available from Akron and Fort Collins, Colorado, and Cheyenne, Wyoming, all within a 140-km radius of the site.

As the major research site for the US-IBP Grassland Biome, a large amount of data were collected between 1969 and 1978. Although many individual process studies were conducted, the majority of the data collected describe the standing crop of biomass at all trophic levels and changes through time. As such, the data should be useful background information for future studies. In addition, similar data are available

from several other grassland sites across the western United States. Most of these data were keypunched, screened, summarized and put on tape. While a few tapes have degraded with age to the point of not being readable, an effort is being made to retrieve the core sets and store them on new tapes. The types of data collected are too numerous to list--more than 80 data formats were designed and used on standardized sets alone. Various descriptive information such as maps also exist; some 111 theses and dissertations, research, and over 200 publications came out of the Grassland Biome.

2. Data Management System

2.1 Goals

Three major purposes may be ascribed to a data management system. (1) Software development costs should be reduced by having general purpose utilities that can be used for several types of data; this commonality should also reduce retraining problems with changes in personnel. (2) By making centralized computer handling of data easier than other ways, screening of data on a timely basis can be encouraged and data quality and availability increased. (3) Once data are made available and documented, they can be more easily safeguarded against both direct loss and loss of interpretability as scientific personnel change.

2.2 Resources

Personnel currently available include two professional programmers with biological backgrounds/interests, a former technician learning programming, and a keypuncher. These personnel are available from an interproject pool for work or consultation on an hourly charge basis.

Funding

Approximately 1 percent of the CPER/LTER budget has been allocated to data management. That is a projection and will be adjusted as required during each funding year.

Facilities

The CSU Computer Center has two Control Data Corporation Cyber 720 mainframes (one for funded projects, one for class and administrative use) and a Cyber 205 supercomputer. Software available includes several statistical packages (SPSS, BMD, Minitab), graphics packages (Tektronix, DISSPLA), and database managers (System 2000, SIR, RIM5, SEED). The most commonly used language for us is Fortran (version 66 with a strong move towards 77), although PASCAL and others are available under the NOS operating system. Simple text editing is available interactively. The LTER has an IBM Personal Computer with BASIC and word-processing (EASYWRITER) capabilities, which can also act as a terminal to another system.

The NREL has recently obtained a Digital Equipment Corporation (DEC) VAX 11/730 because the CSU Cybers are so heavily overloaded and not

designed well for interactive use. A floating-point accelerator (high-speed math processor) and an expansion cabinet are included. It has a card reader (for historic purposes), a 1600 bpi 9-track streaming tape drive (Cipher with National Semiconductor Controler designed primarily for disk backup), and initially one 120-megabyte fixed media disk and one 10-megabyte removable pack drive (from DEC). In the near future two more 160 megabyte drives are anticipated. A high-speed electrostatic printer/plotter (Versatec) is on line. There will be 24 ports (8 DEC, 16 Emulex) available to attach terminals, etc., although we do not expect the system to be fast enough to support more than eight users at a time. There will be three dial-up lines, and the system will have an autodialer so it can dial another computer system (Emulex equipment). Several dumb terminals are available, as well as five colorgraphics (DEC Gigi) terminals, and a Tektronix 4054 vector graphics system with a digitizing tablet and pen plotter. The IBM PC should work easily as a terminal to the VAX. There will also be a batch link to pass files to and from the CSU Cybers. We will also be developing a link to the NREL word-processing system (IBM OS/6 and Mag Card).

We are running the UNIX (Bell Labs) operating system with Berkeley (4.1 bsd or 4.2 bsd) enhancements. We will initially have no statistical packages (we're looking at Minitab) and simple graphics packages. The INGRESS and probably RIM5 database systems are available. Fortran (version 77) will be used most, although C and Pascal will be common and others (APL, etc.) available. Good editors and many utilities come with the system.

2.3 Data Management

2.3.1 Administration

Goals and policies will be set by the LTER P.I., with advice from the steering committee and database manager (one of the programmers). Implementation will be tempered by software/hardware available and personnel sharing with other projects. All data will be encouraged to be computerized or at least indexed by computer (maps, literature), and data sharing will be encouraged. A team approach has been used here for some time, so this may be less of a problem than for other sites. We attempt to form a team including researchers and data managers: researchers are responsible for data quality, documentation, and timeliness; data managers for ease of use (data entry, screening, extraction, combining with other data) and security (both from unauthorized access and inadvertent loss).

2.3.2 Information System

The development of an information system in the computer sense has been delayed, waiting for the VAX. Most of the Grassland Biome data were catalogued and a major indexing/documentation effort was well under way when funding terminated. Both of the programmers and some of the researchers were here during the Biome, and an effort is under way to salvage all that we can (most Biome tapes were written on 7-track drives and need conversion to 9-track, for example). An interactive directory will be developed and kept up to date for new data. Historic data will

be kept only on tape, but current data will be kept on line, probably on a disk pack (10-megabyte removable). Data will be indexed by project (e.g., LTER), study site, data type (plant, insect, meteorological, etc.), treatment, date, investigator, taxonomy, etc. Location, availability, source of documentation, computer formats, related software, summaries of data, related data sets, related publications will be listed where possible. To date, all such information has been gathered by programmers and/or technicians. It is hoped that with easier to use, more visible systems that more of this will be done by researchers.

2.3.3 Data Base Management

Historically, all computerized data have been handled through a computer services group (programmers, keypunchers) so that records were maintained by a group under one manager (or programmer). No usage histories were maintained directly. Within the Grassland Biome, little distinction was made between primary and secondary users except that as requests were received for help, priorities were established on who got served first essentially on a case-by-case basis by the manager in consultation with the Biome director or assistant director. Neither data nor the computer system were well enough documented for someone outside the service group to access them without help. Many of the data were collected by technicians for general purpose use, so the proprietary question rarely arose and was handled individually when it did.

For most data, field collection forms were designed by the computer services group in consultation with researchers, and special screening programs were written for each data type. As data were collected, the forms were sent to the central services group for keypunching and first-pass analysis--typically means and standard errors, as well as machine-detectable data errors. Only batch processing via cards, with magnetic tapes for storage, were available. An attempt was made to process data and return summaries to the field sites within two weeks to provide feedback before the next sampling date. After initial startup on each data type/sampling scheme, this goal was usually met. At the end of each field season, a second set of summaries was run for the full season, incorprating data corrections.

As these summaries were used by investigators at each site and by the "integrators" at the Biome headquarters, another series of corrections were often made. Major sets used by several people were thus "cleaned" far better than one-use sets.

For current data for the LTER, many things have changed. There is insufficient funding for a large single-purpose services group; although the concept remains, it is much smaller and serves several projects rather than one. Computer systems are now easier to use and accessible interactively from many places. Thus, emphasis will have to be changed to "consulting" rather than "doing for"; to computer systems for directories, documentation, data storage rather than interfacing through a group of people; and probably to data input and screening by researchers. We believe the same experience on "the sooner and the more the data are used" (this includes use by more people), the better the data will be.

Access authorization may become a problem but is addressable by our computer systems.

We are not sure yet whether a formal database management system, or just a series of useful utilities/file handlers will be used; both will be available. Each has advantages and disadvantages. All data should pass through screening procedures to help with error detection; simple graphics as well as reasonability checks are often the most useful. However, people looking at and trying to use the data are the more important final test. The LTER data entry will take several forms: batch entry through keypunching or the equivalent, automated systems, and interactive input. The INGRES system has a screen-oriented input front end; it is possible that it will be useable alone without entering directly into the database. Updates and corrections will have to be handled differently, depending on data entry methods--decisions await better knowledge of the VAX system. These decisions will be formalized as soon as possible.

2.3.4 Archiving

Where they are used, original data sheets will be filed until space becomes a problem. We are in the process of putting Grassland Biome forms on microfiche because they now fill eight file cabinets and space is critical. Copies of data forms should be maintained in two locations. Where data are kept on computer disk, we plan daily backups for a month, weekly for six months to a year, and monthly "forever" for the computer system as a whole. However, individual projects (LTER) should make their own at logical points. Full tape copies should be kept with at least one level of backup, and preferably at two locations (NREL and CSU Computer Center). Because of our experience with magnetic tapes degrading, we feel that each set of backups should be checked for readability at least every two years and preferably once a year. Backups should all be made in as machine- and software-independent a form as possible. This is very important with the rapid technological changes now being made. We are not philosophically opposed to special-purpose or locally written software but strongly believe that it should be transportable (machine and operating-system independent) and well documented. Any dependencies that take advantage of local capabilities should be modularized (isolated) and extremely well documented (proved by knowledge and review by more than one person).

Once we have our system up and operating, I hope we will have considerable experience/advice and probably software to share with other sites.

CEDAR CREEK LTER

Data Base Management

Site Report

Prepared by K. C. Zinnel
November, 1983

Ecology and Behavioral Biology Department
318 Church St. S.E.
Minneapolis MN 55455

TABLE OF CONTENTS

A. Status of Data Base Management at Cedar Creek

The primary goal of data management at Cedar Creek is to provide centralized control of LTER operational data so that no information is lost or erroneously recorded. Secondarily, all information obtained in association with LTER research, such as graduate student research or visiting scientists, will be archived as part of the long term data set provided proper documentation is included. The data manager is responsible for these ancillary data only after a clean, well documented data file is submitted by the researcher. When appropriate resources are availiable, other pre-existing data bases will be computerized for easier access.

All procedures for obtaining data are documented during the initial experimental design phase. All researchers are required to use the same codes to identify each measured entity such as field, experiment, treatment, date, and plot designation to avoid ambiguity and inconsistancy. Raw data sheets for recording data in the field (Appendix A), are designed in consultation with the data manager. Variables of a dissimilar nature are not included on the same form in order to prevent any delay in completing the initial raw data entry step. Completed raw data sheets are copied, the originals kept at Cedar Creek and the copies sent to campus for data entry. The resident ecologist is responsible for maintaining the original LTER raw data forms in the records room at Cedar Creek; the data manager is reponsible for maintaining the duplicate copies on campus. Individual researchers are responsible for maintaining their own raw data forms and records, until such time as they turn over properly documented copies to the data manager.

After data file definition "forms" are prepared by the data manager and checked for accuracy, data are entered via microcomputer using comercially availiable software which prompts the operator for required data and does data validation. Duplicate records or missing data are caught immediately, providing prompt feedback in case of procedural problems. Once the integrity of the entered data is verified against the raw data sheets, text files containing the verified data sets are uploaded via microcomputer communication software to the University of Minnesota CDC Cyber 730/845 mainframe computer system. After reverifying the raw data to be sure no data were garbled in transmission, FORTRAN programs are used to prepare data and create variables for addition to data base units (system files) created by SPSS 8.0 (Statistical Package for the Social Sciences, Nie et al.). Summary reports of all data in each data base unit are printed, one copy is kept at Cedar Creek, the other on campus. Direct access to data files is prohibited for security reasons but request procedures are established for the data manager to permit copies of data base units to individual accounts and place copies or subsets of data base units within researchers permanent file catalogs.

A private 844 disk pack is used to archive all data base units and the files and procedures used to create each unit. As part of this procedure an update history and error detection log will be created and maintained by the data manager. Information defining the files availiable, the nature of the variables contained in each data base unit and the procedures for selecting a data set for analysis or intersite exchange will be availiable through an interactive documentation facility (WRITEUP) on the Cyber 730/845 system. For the LTER core data, the PI's, resident ecologist, and postdoctoral researchers are responsible for documenting the data collection procedures and the nature of the data obtained. The data manager is responsible for updating and maintaining the data base units after the data sets have completed the quality assurance phase of data entry. The data manager, under the direction of the principal investigators, coordinates the tabulation, documentation and dispersal of appropriate data sets and/or data base units.

B. LTER Core Data

Unless otherwise indicated, all data sets described below are replicated in 4 fields; 10, 25 and 50 years since cultivation and an undisturbed native oak savanna. The keywords designating each data set are underlined.

Undisturbed Fenced Microplots Nitrogen gradient experiments with 9 treatment levels on 4 X 4 m plots, 6 replicates. Above and below ground fencing excludes mammalian herbivores. Variables include soil chemistries(nitrogen, potassium, calcium, pH, magnesium, and % organic matter), g/m^2 biomass estimates of occurrence of individual plant species, proportional abundance of individual species, species richness, total biomass, diversity, eveness, small mammal densities, and insect densities.

Disturbed Fenced Microplots Equivalent in all respects to the previous experiment except that the ground was disked before establishing the plots and beginning nutrient manipulation.

Single Nutrient Fertilization Also within the fenced area excluding mammalian herbivores, each 4 X 1.5 m plot is fertilized with a single essential nutrient: N, P, K, Ca, Mg, $NaSO_4$, trace metals, water and a control. Variables measured are soil chemistry and total above ground biomass at harvest.

Unfenced Undisturbed Macroplots In order to compare effects of nitrogen gradient fertilization experiments over a large area, ten 20 X 50 m macroplots were sampled in 1982 then treated in 1983 with 3 of the same treatments used for microplots. Variables measured are the same as for microplot experiments but in alternate years. No mammals are excluded from macroplots, and all woody plants are individually marked with permanent metal tags.

Deer Exclosures For the 2 youngest fields, 3 controls and 3 treatment areas have been completely enclosed to exclude deer

with all woody plants individually tagged. Data recorded are pre-treatment browsing activity on a plant by plant basis, size and abundance of plants.

Small Mammal Grids In a young (10 years since cultivation) field (E), 3 grids of 4 X 4 m plots with 4 levels of fertilization (16 replicates) were established in 1983. One grid was disked prior to fertilization, and one grid is a totally unmanipulated control. Data are soil chemistries, plant distribution and abundance, small mammal density estimates, gopher mound counts, and grasshopper activity censuses.

Unfenced Microplots In fields A, B, and C, 48 microplots (each 4 X 4 m) with 3 levels of fertilizer (16 replicates) were established in 1983. These plots are unfenced and will be sampled more destructively than fenced microplots for the purpose of looking at below ground biomass, population and physiological responses of herbivores, and nitrogen fixation by grasses.

Burn Microplots Modeled after Konsa Prairie burning experiments, 8 X 8 m plots in a 25 year old field were established with burning to begin in 1984. Data comparable to the fenced microplot data were collected by sampling vegetation in 1983. Burning treatments (4 replicates) are control (none), every year, every 2nd year, and every 4th year.

Old Field Survey Permanent 25 m long transects, 4 per field were established in 23 old fields ranging from 1 to 50 years since cultivation. Data collected were % cover of individual species, gopher mound densities, and soil chemistry from 25 quadrats (1 X 1/2 m) along each transect.

Blowdown Plots Violent weather caused extensive damage by blowdowns of mature trees in forested areas of Cedar Creek. Permanent quadrats with sampling transects have been established to monitor responses of understory and overstory vegetation to major climactic disturbance.

Oak Savanna Burn Compartments Permanent 25 X 40 m quadrats were established in 1983 within 12 compartments ranging from 2.6 to 28.6 ha. These compartments have been burned at intervals of 1 to 8 years beginning in1964. Herbaceous, shrub, and tree cover plus selected soil parameters will be measured.

COMPUTER HARDWARE AND SOFTWARE

A. Data Entry

Three XEROX 820 microcomputers are used to enter data using DATASTAR software by MicroPro. Researchers are requested to consult with the data manager before designing raw data recording sheets so that they are aware of the features and limitations of this package. For each type of raw data recording sheet, a corresponding DATASTAR "form" is created by the data manager. This defines the variables to be captured and creates a template for constructing data files. Data entry technicians fill in blank lines on the screen with matching data items from the raw data sheets. Features of the software found to be useful include range checking, formatting control, automatic field copying to eliminate entering redundant data, index sequential file structure allows corrections without searching entire data base, and the ability to access other DATASTAR files to look-up allowable codes for validity checking.

B. Communications

Communication with the University of Minnesota mainframe computers is accomplished by using the Xerox 820's with a D.C. Hayes Smartmodem connected to a telephone jack. Software (COM) purchased from the University Computer Center (UCC) MicroGroup handles normal interactive processing as well as uploading and downloading files from the mainframes to the micros. A microcomputer research lab has been established and stocked with machines most frequently used by the University of Minnesota computing community. Consulting support on software packages for APPLE II, TERAK, XEROX 820, IBM-PC and Zenith Z-100 are currently availiable. In addition to COM, the communications package, UCC sells graphics terminal emulators, utility packages, and text processing tools.

C. Mainframe Computers

University Computer Center (UCC) facilities used by the LTER project include a cluster of 3 CDC Cybers and a CRAY I Class VI supercomputer. For data management, we are primarily using a dual central processor CYBER 730 which is due to be replaced by a Cyber 845 in December, 1983. Currently, data analysis is also done on the Cyber 730 but we plan to use numerical analysis techniques for data collected in 1983 which require matrix manipulations and memory sizes available only on the CRAY I. A wide variety of software packages are supported by UCC; those used to manage LTER data are FORTRAN V, SPSS, ARCHIVE, WRITEUP, and several Permanent File Utilities.

DATA MANAGEMENT PROTOCOL

A. Quality Assurance

All core LTER data are subjected to the following procedures in order to ensure that data are collected correctly, consistently and without ambiguity. To provide a framework for organizing data collected, we use a concept of case identifier or CASEID. Each CASEID refers to a particular plot of ground at Cedar Creek. Identifying information associated with each CASEID include LTER field designator (A, B, C, D, or E), experiment type (1-12), plot number, subplot (if macroplot), year initiated, area, length, width, manipulation imposed, treatment level, blocking factors, and Cedar Creek old field inventory number. Thus, each measurement such as soil nitrogen ppm, light attenuation, weight of a particular species of plant, or observations such as grasshopper or gopher mound counts must be recorded with the appropriate identifying information. Using a DATASTAR CASEID database, the appropriate CASEID is supplied if the identifying information is valid. CASEID ultimately defines the position of that datum as an SPSS variable.

Another DATASTAR database (TAXONKEY) is used for verifying plant species codes during data entry. Encoding dry weights of individual species sampled for each plot would require repetitive entry of taxonomic names except for the ability to "lookup" taxonomic names for sight verification. Raw data sheets used for 1983 plant data (Appendix A) were created by using 1982 data to establish which species were abundant (≥1%) for each field. A set of 4 forms were generated for each field; abundant forbs, abundant non-forbs, rare forbs and rare non-forbs sorted in alphabetical order by taxonomic name. Abundant species were fewer in number but were found most frequently. When rare species occurred, more raw data sheets were required, but overall much time was saved by not having to search through a list of 300+ names to make an entry. Also fewer data items were missed during data entry.

Raw data sheets are copied as soon as they are completed. The originals stay at Cedar Creek, copies are sent into campus for initial entry via DATASTAR form definition files. After completeing this step, all data are printed and checked for errors against the raw data sheets. Errors are corrected using built-in DATATSTAR record retrieval features such as key search, masks, or index order to recall and fix erroneous data. During the data entry and verification process, a back-up copy of each disk is made at the end of each entry session. After an error-free verification, the data manager is notified that the data set is ready for uploading. For security reasons, this procedure is done only by the data manager. A hard-copy printout of the uploaded data set is then made and verified against the hard-copy printout from the final error free DATASTAR disk. The duplicate

DATASTAR disk and hard-copy printout are then stored at Cedar Creek. The original DATASTAR disk and hard-copy printout after uploading remain on campus.

B. Data Base Access

We have found that trying to include all possible variables in an SPSS system file creates an uneconomical situation as many missing value indicators are required to fill out the matrix. Instead, we have decided to organise our data base into units based on comparable experimental situations. For example, all macroplot data would be one unit because the sampling regime is different from microplots. Old field survey data are another example. These units are prepared and updated by the data base manager. FORTRAN programs are used to sort and calculate variables desired in addition to original measurements. In the case of the plant data; total biomas, species richness, diversity, other indices and proportional abundance of individual species are calculated on a plot by plot basis. Files are constructed with variables or missing value indicators in the appropriate cells to match the structure of each SPSS system file unit. A final SPSS add variables procedure and save archive file job is required to complete the "build". SPSS procedures summarizing all new data are then run and compared to test cases selected at random. Access to these units by LTER PI's, researchers and staff is given by the data manager using permanent file utilities to permit read-only access from other user accounts to the LTER master account where the files reside. Copies of any files can be placed within other user catalogs by the data manager if requested. Access to the main LTER account is restricted for security reasons to the Data Manager and the PI's.

C. Data Archiving

All files used in a build (raw data, intermediate data, FORTRAN programs and SPSS procedures) to create or update a data base unit are collected into a single master file using the LOADPF/DUMPPF utility. These files plus the resultant data bases are saved on the LTER private disk pack. No other user accounts are permitted to use this 200 Mbyte capacity device. Copies of the build files are also stored on magnetic tape using the ARCHIVE utility which automatically stores the files in a format that mimics permanent files and allows reloading files as permanent files with automatic conversion. All raw data, intermediate files, and procedures are maintained as text files; only the data base units are binary, i.e. SPSS system files.

D. Intersite Exchange

Researchers desiring access to Cedar Creek LTER data must make a written request defining those variables or data sets of interest. The preferred medium of exchange is magnetic tape, supplied by the requesting site. Text files of the appropriate variables from one or more data base units will be prepared by

the data manager. The initial file on the tape will be a text file defining the information contained in each file and the variables contained in each column. The tape will be written using magnetic tape options considered the industry standard for information interchange; 9-track, 1600 PE (or 800 NRZI) B.P.I., unlabeled, blocked BCD, EBCDIC (or ASCII) encoded. Computer manufacturers maintain consistency in defining these formats for this purpose. Sites desiring to use interactive access (and long distance telephone charges) may obtain user numbers by calling UCC accounts office (612) 373-4548 to request one, or write directly to University Computer Center, 227 Experimental Engineering, University of Minnesota, 208 Union Street SE., Minneapolis, MN 55455. In this case, copies of desired data files will be placed within the requesting user catalog. A usage history log will be maintained by the data manager, and all users are required to acknowledge the source of data used in any publication. A suggested form of acknowledgement is: "data on total biomass as a function of nutrient manipulation were obtaintained from the Cedar Creek LTER project, NSF grant # DEB-8114302, awarded to G.D. Tilman and J.R. Tester."

E. Data Documentation

The purpose, methods, and all other relevant information on all research done as part of the Cedar Creek LTER project will be described in a Data Catalog using an interactive documentation utility (WRITEUP) availiable on the UCC Cyber 730/845. Format will consist of an introductory section which describes the structure of the documentation file and an index which lists the keywords needed to access documentation for different data sets. The general structure of the data documentation file will follow that described for developing data documents in the Long Term Ecological Research Data Management Workshop report from Urbana-Champaign, Illinois, 1982. Described under the sub-heading "LTER Core Data Sets", current research requires 10 sections; more sections will be added when completed. As work progresses, new information will be added as subsections of the appropriate section defined by index keywords. Each section of this Data Catalog will describe a particular type of experiment or series of observations comprising each data set. We will attempt to show the actual layout of the experimental plots in each field, describe the experimental treatment associated with each CASEID, detail exactly what sampling regime was used for each experiment, and provide additional pertinent notes, such as whether conditions or techniques changed, in chronological order. Also associated with each section will be information describing the procedures for accessing the current data base unit (SPSS system file) corresponding to that set of documentation and listing variable labels assigned to the data contained in each data base unit.

Each datum collected is considered to be a variable (column) associated with a particular CASEID (row). All variables are named in an internally consistant manner using alpha and numeric character codes. For example, P83W103 is an acronymn for Plant

abundance 1983 Weight of species 103. All plant variable names begin with P, chemistries with C, indices with I, etc. There are distinct identification codes for the scientific names of all taxa or other property measured such as nitrogen ppm. Date and sample period are also indicated by the variable name.

SUMMARY

A. Achievements

Data management for the Cedar Creek Long Term Ecological Research project has achieved four goals. Data collection procedures assure that data are recorded correctly, consistently and without ambiguity. Data entry and data verification standards assure that data integrity is maintained throughout the data base creation/addition process. All raw data, data used for analyses and procedures to manipulate data are archived and backed-up for future use. Documentation of all relevant data sets and complete descriptions of experimental conditions, variables, measurements, and observations are being prepared to facilitate intersite exchange of information.

B. Problems

As pointed out at the 1982 LTER Data Management workshop at Urbana-Champagne, Illinois; the primary role of the data manager is to facilitate the availiability of data generated by research at LTER sites. However, in reality each data base manager responds mainly to the needs and priorities of the individual P.I.'s and is limited by the amount of resources allocated to data management. Because of the ambitious nature of LTER objectives, the position of data manager is more than 1/4 time or even 1/2 time, it is basically a full-time responsibility. In dealing with problem situations, such as locating missing data items or dealing with hardware problems, immeadiate resolution is necessary. Adding new experiments, changing sampling regimes or variables measured, creates time consuming tasks for the data manager in that previously used forms and programs need to be changed accordingly. Future usage of LTER data require an initial investment of resources and commitment to data management support. Without full-time professional data processing support staff, data management will not be consistent or effective.

APPENDIX A

Examples of raw data sheets corresponding to DATASTAR full screen data entry software.

A. Plant distribution and abundance.

B. Percent cover estimates for Old Field Survey.

C. Gopher mound density.

~~FIELD~~ C PLANT DATA FORM (NON-FORB SPECIES) EXAMPLE A

PLOT C-1-2C DATE 17 Aug. 83

		PRESENT	MIN	MAX	MODE	WT.
100	AGROPYRON REPENS					
102	ANDROPOGON GERARDI					
103	ANDROPOGON SCOPARIUS	✓	18.0	69.5	45.0	57.60
200	CAREX SP.					
106	CENCHRUS LONGISPINUS					
2	CLADONIA & LICHENS					
313	CORYLUS AMERICANUS					
230	CYPERUS FILICULMIS					
231	CYPERUS SP.					
1	EQUISETUM LAEVIGATUM					
107	ERAGROSTIS SPECTABILIS					
199	MISC. GRASSES					
399	MISC. WOODY PLANTS					
99	MISCELLANEOUS LITTER	✓				132.36
98	MISCELLANEOUS SP.					
109	PANICUM CAPILLARE					
111	PANICUM OLIGOSANTHES					
112	PANICUM PERLONGUM					
113	PANICUM PRAECOCIOUS					
117	POA PRATENSIS					
132	POA SP.	✓	6.0	70.0	30.0	21.88
304	ROSA ARKANSANA					
305	RUBUS SP.					
118	SETARIA LUTESCENS (GLAUCA)					
120	SORGHASTRUM NUTANS					
122	STIPA SPARTEA					

Clipped by: Sue

Sorted by: Bruce, Stacy, Kathryn, Dan, Nancy

EXAMPLE B

Old Field Census Data Sheet

Date: 83/07/____ Field #:_____

Transect #:_____ (G,R,W,Y) Plot #:_____ (1-25)

Taxon Key	Species	% cover (or Trace)
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

O. biennis plants= _____ # in flower= _____

bare ground_____ _____

litter_____ _____

sum= _____%

trace % = 100 - sum = _____%

number of trace species = n = _____

% for each trace species = (trace %)/n = _____%

EXAMPLE C

Gopher Mound Data Sheet for Old Field Census

Date: 83/~~07~~ 08/02 Field #: 76

Bruce

Tagged Stakes

Y	Quadrat 6 # mounds = 19	Quadrat 5 # mounds = 18
W		
	Quadrat 4 # mounds = 23	Quadrat 3 # mounds = 55
R		
	Quadrat 2 # mounds = 18	Quadrat 1 # mounds = 39
G		

Yellow

White

Red

Green

H.J. Andrews Data Management Update
November 1983

I. Data acquisition:

380 data sets are in the data bank, corresponding to about 1000 separate datafiles and 70 X 10^6 characters of data.

II. Data documentation:

Ten out of twelve LTER studies have complete documentation thus far in the progress of these studies (data abstracts, datafiles cataloged, data formatting information recorded).

Abstracts have been computerized (using the FAMULUS system) for all LTER studies for which we have completed abstracts. In addition we have computerized the abstracts of some other active or recent studies (total = 60).

We have just implemented a system for computerizing the data formatting information. Only a few LTER studies have data formatting information computerized at this time.

III. Information retrieval

During the last six months we implemented a microcomputer database management system (INFOSTAR, from Micropro, Inc) which is being used for forms data entry and verification, sorting merging and data restructuring; as well as for maintaining mailing lists, the data directory and data formatting information. Through the INFOSTAR system we have been able to implement a 'on-line' data catalog for data bank users at several key microcomputer locations. The INFOSTAR system is composed of DATASTAR, SUPERSORT, and REPORTSTAR.

The INFOSTAR package largely replaces our previous data cataloging system which was based on in-house software written in CBASIC, Pascal MT+, and FORTRAN5. We are still using FAMULUS to retrieve data documentation information on the CYBER mainframe computer.

IV. Hardware and software

a. Hardware.

During 1983 we made several major hardware acquisitions with funds coming from a variety of sources. Two low cost 8-bit microcomputer systems were acquired for data entry. Three CompuPro 8-bit/16-bit switchable processor microcomputer systems were purchased for data editing, analysis, graphics, and communication with the OSU CYBER computer and the University of Oregon IBM mainframe computer. In addition a Lomas 16 bit multiple user computer system was acquired with a hard disk (this system is not yet operational).

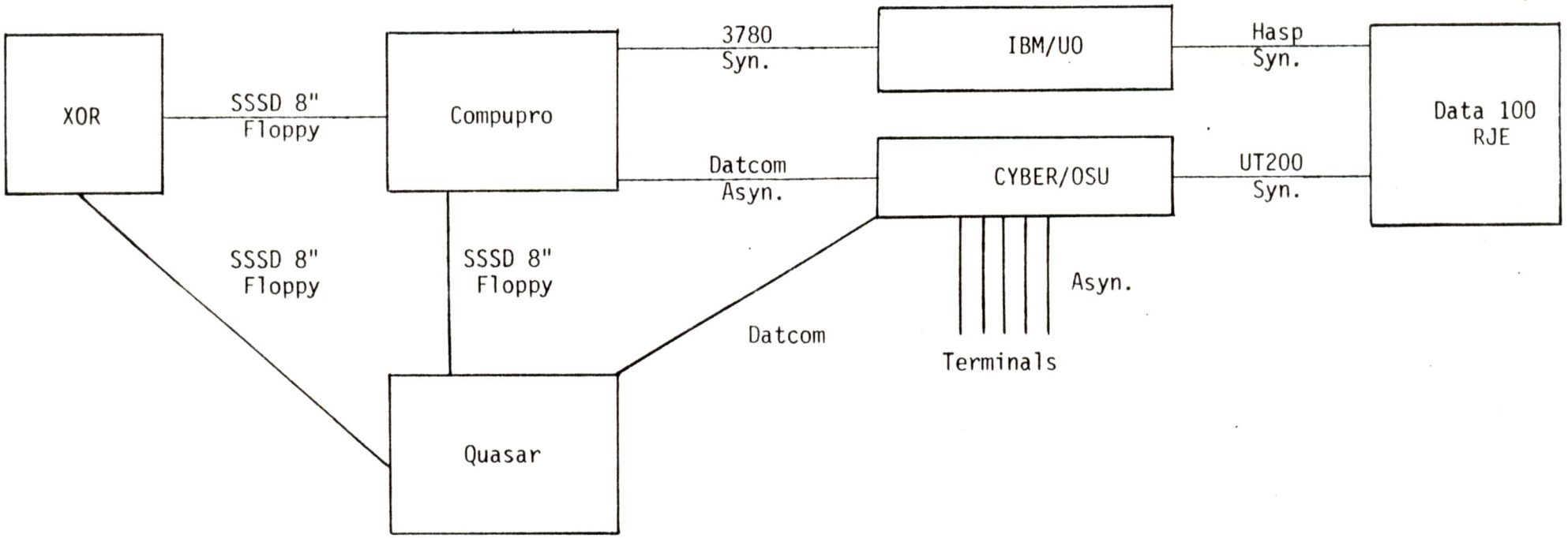
We also acquired some new peripherals which gave our computer network some new capabilities: high resolution color

graphics high speed printing, and word processing. We now have 2 word processing printers, a monochrome graphics terminal and a color graphics terminal. We also acquired a 300 line per minute printer which is shared by the CompuPro microcomputer network.

b. Software.

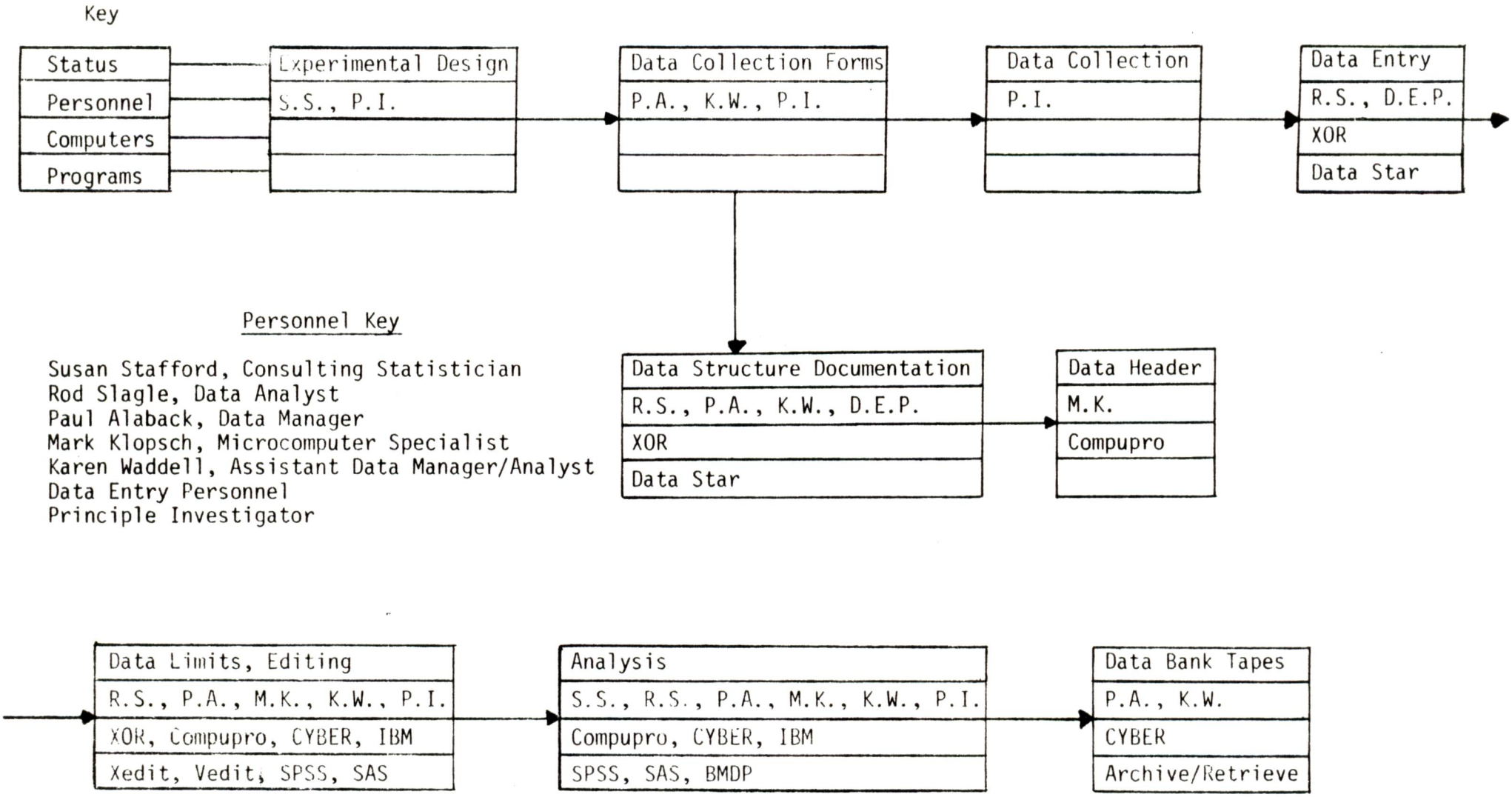
We considerably expanded our software base during 1983. Most important was the implementation of a OSU computer center data communications package providing error checking for the transferal of data files between the microcomputers and the mainframe. We also acquired a 3780 IBM protocol communications package which allows our microcomputers to communicate with the University of Oregon IBM computer directly. The prime reason for developing this communication was to access the powerful SAS statistical package (not available on CYBER computers). In addition we expanded our microcomputer software base to include two versions of BASIC, Pacsal, Fortran4, Ada, two database management systems (DATAFLEX and INFOSTAR) and a graphics language (for the Vectrix). For more details see handout on software and hardware.

FSDB COMPUTER NETWORK



Micro Machines	# In Use	Operating System	Disk Format	Functions
XOR	2	CP/M 80	SSDD 8"	Data Entry Word Processing Documentation
Quasar	3	CP/M 80	DSDD 8"	Data Entry Word Processing Programming Digitizing
Compupro	2 1	CP/M 80 CP/M 86	DSDD 8"	Programming Graphics Communications High Speed Printing

QSG DATA FLOW CHART



Baruch Institute
UNIV S. Carolina

DATA MANAGEMENT UPDATE

1. INTRODUCTION

Developments in research data management have proceeded rapidly since January 1983 when two part-time data management positions were created. Supervision of data entry, quality assurance, archival, and programming are divided between the two data managers. Marvin Marozas manages the physical and chemical data sets, and Bill Michener manages the populations-related data sets. A summary of accomplishments, difficulties encountered, and future goals within each category is described below, followed by an update of hardware and software enhancements, and accomplishments and planning within related areas.

2. PHYSICAL AND CHEMICAL DATA SETS

Chemical data sets for the LTER project are currently subdivided into two major categories: 1) Daily Water Sample (DWS) and 2) Interstitial Sediment Chemistry (TSC). Three others: 1) Forest Nutrient Transformation (FNT), 2) Fresh Water Stream Analysis (FWS), and 3) Leaf Decomposition Study (LDS) will be added within the next six months. The DWS data management scheme has been refined to the point that several months of chemistry data can be entered into the system by use of full screen management programs, verified, corrected to concentrations, automated reports written, and data archived, all within three man days. Two data sets which are locally managed by different individuals (Chlorophyll and ATP), will be added to the DWS

data set for final archiving procedures. A similar scheme has been developed for the ISC data. The first sampling period, November 1981, has been used to test the scheme with good results. By September, the first year of ISC data (November 1981 through August 1982) should be worked up and archived.

The meteorological data being collected by the CLIMATRONICS system is currently in two formats. One is digitized on cassette tape, and the other is hourly averages block printed in code on standard computer printout. Neither of the present formats is useful for the researcher desiring any type of time series analysis of the data. The purchase of a DEC microcomputer along with some additional hardware/software items awaiting purchase will allow transfer of the meteorological data to the host system (i.e. the University's Computer Services Division). Once on the host system, the data will be organized and archived. Interested researchers will then be able to access shared copies for subsequent manipulations. In the future, a program will be devised that will provide a synopsis of monthly meteorological data that can be sent to interested researchers on a regular basis.

3. POPULATION-RELATED DATA SETS

Full screen management entry programs have been completed for all population-related data sets except "meiobenthos" where data entry procedures had been developed previously. Three technicians were hired for the summer to complete entry of all raw data collected to date. Currently, all population data has been entered and checked for key-

punch errors. Future data entry will be accomplished by the technician most closely allied to the project (i.e. the technician who collects and analyzes zooplankton samples will also enter the data and check for key-punch errors). All routine output programs such as those providing printouts of raw data in customized report formats have been completed. Programs developed or under development include internal documentation.

Population data has been entered in its rawest form (counts per sample) and temporarily archived in mass storage system. Copies of the data exist in several locations. Array-processing programs which translate raw data into standardized form (i.e. numbers of organisms per cubic meter) are currently under development. Array-processing programs and programs which provide exploratory statistical analysis of the data are expected to be completed within the next 2-3 months. Output from these programs will be archived as secondary files to the original.

Generally, data entry and program development has proceeded much faster than anticipated. Future plans (1984) include the addition of several related data sets to the LTER data base and the development of programs which provide updated exploratory statistical and graphical analyses in report format on a routine basis.

4. THE BARUCH DATA BASE AND DATA CATALOGUE

Presently, all LTER data sets being dealt with by the data management group are being temporarily archived on CSD's Mass Storage

System and backed up on either hard disk or magnetic tape. Our most important long term goal is to see that LTER data, along with all of the Baruch Institute's data, is properly archived with precise documentation in a data base management system (DBMS) which will best serve the Institute and the researcher. The main criteria which the Baruch Data Base system must meet will be: 1) That it is dynamic to allow for the addition of new data types quickly and easily, 2) simple to use and maintain, 3) safe, and 4) relatively inexpensive. We have done a great deal of research into the systems supported by CSD and have been working closely with Ron Bonnel's graduate student, Betty Whitaker, in designing a system which will meet our criteria.

Unlike a business DBMS, a research DBMS must have thoroughly documented data for it to be useful to secondary users. The guidelines for the documentation procedures have been laid down by the LTER Data Management Workshop (University of Illinois, November 1982). We have found that the documentation of data sets is one of the most tedious and labor intensive aspects of data management, and we are currently exploring the possibility of developing a computerized documentation procedure which should decrease time alloted for this task. The compiled documentation which will be put in the form of a data catalogue will be stored in the Baruch Data Base as an integral part of the system.

5. HARDWARE/SOFTWARE

We have recently acquired a new Courier CRT (now a total of 3) and DEC Professional 350 computer. Pro/Basic and a word-processing package have been initialized on the Professional 350. Statistical

programs have been developed and added to the Professional allowing for some in-house analysis of data. Terminal emulation and communications software are expected to be purchased in the near future. These features will support communications with The University of South Carolina's Computer Services Division (CSD), the CLIMATRONICS meteorological system, and computers at other LTER sites. The expected acquisition of a Hewlett-Packard color graphics plotter will considerably enhance in-house data analysis capabilities and decrease our charges for use of equipment from CSD. CSD recently acquired VM/CMS (Virtual Machine/Conversational Monitor System) as a replacement to VSPC (Virtual Storage and Personal Computing). VM/CMS will alleviate current problems with limitations in file size and enhance interactive capabilities with other software products (i.e. Statistical Analysis System). However, modifications will need to be made to some existing VSPC programs for execution under VM/CMS. Because of the imposed deadline by CSD (VSPC will be eliminated from the system December 31, 1983), program and file conversion will assume a high priority on the data management task list. We are currently setting up VM/CMS accounts for our users and plan to have all VSPC files converted to VM/CMS files within the next three months.

6. Related Developments

As data management capability increases and new hardware and software products are purchased, the need for education of staff members concurrently increases. In order to provide this service we have recently begun to teach classes and conduct seminars at the laboratory.

To date, instructional classes in the use of microcomputers, CRT's, and VSPC have been conducted as well as seminars on the various aspects of data management. These seminars and classes have been well received. Consequently, classes detailing use of VM/CMS, SAS (Statistical Analysis System), and the DEC Professional 350 are scheduled for the fall.

Colorado

PREFACE

The following report contains data summaries and accompanying text from projects involved in the University of Colorado Long-Term Ecological Research program (CULTER). The report is intended for internal circulation within the Long-Term Ecological Research (LTER) network, so that relevant data are readily available to all LTER scientists. This data report is not intended to be a citable reference, because data and interpretations are preliminary and later may be updated or altered. If an individual wishes to cite data or interpretations contained in this report, the senior author should be contacted for updated information by using the form on page v and the material should be cited as a personal communication. Use of the form will help CULTER evaluate the importance and extent of data circulation.

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Additional copies of this data report are available at cost from: LTER Coordinator, Institute of Arctic and Alpine Research (INSTAAR), University of Colorado, Campus Box 450, Boulder, Colorado 80309.

CULTER encourages collaborative research at the Niwot Ridge site and interested researchers are urged to contact the principal investigator: Dr. Patrick J. Webber, Director, INSTAAR, University of Colorado, Campus Box 450, Boulder, Colorado, 80309. Investigators interested in collaborative research at other LTER sites also may request the booklet, *Long-Term Ecological Research in the United States: A Network of Research Sites.*

UNIVERSITY OF COLORADO LONG-TERM ECOLOGICAL RESEARCH PROGRAM (CULTER)

Personal communication about data from CULTER Data Report

From:

Requestor_____ Date _____

Address_____ Affiliation_____

_____ _____

_____ _____

To:

Senior Author _____

Data Report Number_____

Title_____

(1) Information to be cited (brief description of data or interpretation for which personal communication is desired; give details on other side if necessary).

(2) Proposed use of personal communication (e.g., title of journal to which paper will be submitted).

Updated citable information has been provided by:

_____ _____

Signature of senior author Date

Photostat copies of this completed form should be sent by the data report senior author to: 1) requestor, 2) Dr. Patrick J. Webber (Institute of Arctic and Alpine Research, Campus Box 450, University of Colorado, Boulder, 80309). Original should be retained by the senior author.

Acknowledgments: The requestor should acknowledge the authors and the CULTER Program, which is partially funded through NSF grant DEB 8012095. Other grants may be applicable to a specific CULTER project.

Resulting Documents or Reports: Please supply CULTER (Dr. Patrick J. Webber) with a notification of talks given or two copies of any written material resulting from this request.

University of Colorado Long-Term Ecological Research Data Reports

Halfpenny, J.C., Ingraham, K.P., and Adams, J.A., 1983: Working atlas for the Saddle, Niwot Ridge, Front Range, Colorado. University of Colorado Long-Term Ecological Research Data Report, 83/1. 24 pp.

Caine, N., Furbish, D.J., and Moses, T.A., 1983: Stream-flow data from the Green Lakes Valley, 1981-1982. University of Colorado Long-Term Ecological Research Data Report, 83/2. 56 pp.

Furbish, D.J., Brendecke, C.M., Caine, N., and Moses, T.A., 1983: Ablation, evaporation, and evapotranspiration in the Green Lakes Valley, 1981-1982. University of Colorado Long-Term Ecological Research Data Report, 83/3. 16 pp.

Short, S.K., Nichols, H., Briggs, W., Elias, S.A., and Andrews, J.T., 1983: Paleoenvironmental studies in the Colorado Front Range. University of Colorado Long-Term Ecological Research Data Report, 83/4. 38 pp.

Mancinelli, R.L. and Shulls, W.A., 1983: Population dynamics of soil bacteria from Niwot Ridge, 1981-1982. University of Colorado Long-Term Ecological Research Data Report, 83/5. 13 pp.

Flock, J.W., 1983: Changes in plant communities between 1972 and 1982 on Niwot Ridge. University of Colorado Long-Term Ecological Research Data Report, 83/6. 38 pp.

Caine, N., Thurman, E.M., Furbish, D.J., and Moses, T.A., 1983: Water quality data from the Green Lakes Valley, 1981-1982. University of Colorado Long-Term Ecological Research Data Report, 83/7. 47 pp.

Greenland, D.E. and Daly, C., 1983: Microclimate data from the Saddle Area, Niwot Ridge, Summer - 1982. University of Colorado Long-Term Ecological Research Data Report, 83/8. 36 pp.

Halfpenny, J.C. and Pollak, O., 1983: Snow survey data from the Saddle, Niwot Ridge, Colorado, 1981-1982. University of Colorado Long-Term Ecological Research Data Report, 83/9. 53 pp.

Losleben, M.V., 1983: Climatological data from Niwot Ridge, East Slope, Front Range, Colorado, 1970-1982. University of Colorado Long-Term Ecological Research Data Report, 83/10. 193 pp.

Bushnell, J.H., Butler, N.M., Haney, B.G., and Wahle, B.M., 1983: Benthos data from the Green Lakes Valley flowage system, 1981-1982. University of Colorado Long-Term Ecological Research Data Report, 83/11. Submitted.

Bushnell, J.H., Butler, N.M., Haney, B.G., and Wahle, B.M., 1983: Plankton data from the Green Lakes Valley flowage system. University of Colorado Long-Term Ecological Research Data Report, 83/12. Submitted.

Windell, J.T., 1983: Brook trout (Salvelinus fontinalis) population status in the Green Lakes Valley aquatic ecosystem. University of Colorado Long-Term Ecological Research Data Report, 83/13. Submitted.

Jones, W. Kamani, G., Howard, V.E., and Komarkova, V., 1983: Natural recovery of plant communities on disturbances in the Niwot Ridge/Green Lakes Valley Area. University of Colorado Long-Term Ecological Research Data Report, 83/14. Submitted.

Keigley, R.B. and Webber, P.J., 1983: The response of Kobresia vegetation and soil to irrigation, nutrient supplements, acid rain, shade, and increased snowpack - 1982. University of Colorado Long-Term Ecological Research Data Report, 83/15. Submitted.

Pollak, O. and Webber, P.J., 1983: Soil moisture, decomposition rates, above-ground primary productivity and standing crop, and phenology data fr six vegetation noda, Niwot Ridge Saddle -1982. University of Colorado Long-Term Ecological Research Data Report, 83/16. Submitted.

Halfpenny, J.C., Golian, S., Ingraham, K.P., Hause, H., Meaney, C., and Southwick, C.H., 1983: Small mammal studies in the Colorado alpine tundra, 1981-1982. University of Colorado Long-Term Ecological Research Data Report, 83/17. Submitted.

Ives, J.D. and Hansen-Bristow, K.J., 1983: Treeline-timberline. University of Colorado Long-Term Ecological Data Report, 83/18. Submitted.

University of Colorado Long-Term Ecological Research Working Papers

Greenland, D.E., 1983: Energy balance instrumentation. University of Colorado Long-Term Ecological Research Working Paper, 83/1. 18 pp.

Greenland, D.E., 1983: The theoretical basis for energy budget measurements on Niwot Ridge. University of Colorado Long-Term Ecological Research Working Paper, 83/2. 23 pp.

Shushan, S., Flock, J.W., Funk, D., and Bonde, E.K., 1983: Plant voucher specimens from Niwot Ridge. University of Colorado Long-Term Ecological Research Working Paper, 83/3. 8 pp.

McNeely, R., 1983: Preliminary physical data on Green Lakes 1-5, Front Range, Colorado. University of Colorado Long-Term Ecological Research Working Paper, 83/4. 34 pp.

Greenland, D.E., 1983: Summer microclimate observations over alpine tundra. University of Colorado Long-Term Ecological Research Working Paper, 83/5. 18 pp.

Halfpenny, J.C., Ingraham, K.P., Adams, J.A., and Peckar, H., 1983: Mountain Research Station: Its Environment and Research. 20 pp.

Halfpenny, J.C. (ed.), 1982: Ecological Studies in the Colorado Alpine: A Festschrift for John W. Marr. University of Colorado, Institute of Arctic and Alpine Research Occasional Paper, 37. 147 pp.

NEW MEXICO STATE UNIVERSITY
JORNADA LTER SITE
November 1983

RESOURCES

The Jornada LTER project supports a one-half time secretary for data keying; other formal data management duties are assumed by Conley, with consultations with other site P.I.'s on an 'as needed' basis. Total computing resources supplies and services budget is approximately $6,000 above the half time position above.

QUALITY ASSURANCE STANDARDS

Individual researchers are responsible for the merits of data sets within their perview. The Data Management Office (DMO) attempts to facilitate data screening and correction by working with researchers who request such help. A series of data-test programs have been written that serve to flag unacceptable data points. The results of these filtering programs are given to cognizant researchers who thus have opportunity to review and revise as required.

ARCHIVING

Each researcher maintains a copy of all field data forms. The DMO additionally maintains notebooks containing all raw forms provided for data entry. Data are entered at a workstation (see beyond) onto 5.25" disks, and then transferred to mag-tape files on a third generation mainframe computer. Duplicate tapes are being maintained, and short term plans (by Feb 1984) include site separation of duplicate tapes. At minimum, three duplicate versions of all data sets exist, two copies of the raw data forms, and one copy of mag-tape form.

DATA ACCESS

A request by the cognizant researcher to the DMO (Conley) results in the desired data set being transferred to an agreed upon location.

DATA ACKNOWLEDGMENT

Data access is arranged by the cognizant researcher; it is assumed that the provider and the providee have agreed on the form of any acknowlegment. It is also expected that appropriate acknowledgment of the NSF/LTER program and the site will be made.

HARDWARE/SOFTWARE CONFIGURATION

HEWLETT-PACKARD MINICOMPUTER: Housed in Population Ecology Laboratory
Wildlife Science, 118 Knox Hall

An active upgrade program is currently in force; the system

described herein should be complete and on-line by February 1984.

HARDWARE:

CPU: Hewlett-Packard 1000/A600+ with 1 Mb memory running HP RTE-A with VC+ Rev B.83.

Disks: a) H-P 9133B 10 Mb Winc. with 3 1/2" microfloppy drive (270Kb).
b) H-P 7914P 132.1 Mb with 16 or 65 Mb cartridge tape device.

Terminals: a) Two H-P 2627A color-graphics display terminals.
b) H-P 2642A workstation with text processing and 5 1/2" floppy disk drive (325 Kb).
c) H-P 2621B editing, programming terminal.

Plotter: H-P 7475 8 1/2 x 11 or 11 x 17 in. 6 pen plotter.

Printers: a) C-Itoh dot matrix as general purpose.
b) Toshiba P1350 dot matrix for graphics and special character printing.

Videocamera: Lang Systems Videoslide (Takes RGB from 2627A to 35 mm).

SOFTWARE:

Operating system: H-P RTE-A with VC+ (Provides 6.8 Mb main memory addressing and 128 Mb virtual memory addressing. This is a multi-user, multi-tasking, real-time system.)

Compilers: a) Pascal/1000
b) FORTRAN 77

Data Base Management: H-P Image/1000 II, a 'relational' DBMS.

Statistics: a) Minitab/1000
b) In house packages by Conley

Graphics: H-P DGL/1000 (2-D) and AGP/1000 (3-D).

Accessory Software:
a) GRAFIT/1000 (A full service conversational high resolution graphics system).
b) SKETCH (A high resolution graphics system for "drawing").
c) TEX/1000. A wordware system by Donald Knuth (TEX83) of Stanford - does technical material, accredited by the American Mathematical Society.
d) JSPEL. Backup spelling system.
e) Locally developed packages as required.

The above Hewlett-Packard 1000 system is new in 1983. Total list value is approximately $80,000, with the NSF/LTER funding about one-third of these costs.

NMSU CAMPUS MAINFRAME

The NMSU Central Computing Facility is being used primarily for data storage (mag tape), and large analyses (via SAS etc.). The current computer is an Amdahl V-5 processor with 4 Mb memory, running IBM's CMS. Software support includes the usual packages expected at a large computing facility; those of most interest to LTER include SAS, SAS/GRAF, PLOT10 (a Techtronix graphics emulator) and BMDP.

MICROCOMPUTERS

A cluster of Northstar Horizons and an Advantage serve for some data entry, data logging, and simple data analyses. This system has the following hardware components:

- Standard Z80A CPU, 64 Kb RAM, Dual quad capacity 5.25"-disk drives, disk controller board, motherboard, two serial and one parallel I/O interfaces, and a power supply.
- Zenith Model Z19 video terminals with a 12 inch (diagonal), P4 Phosphor CRT and an 84-key (plus a 12-key numeric pad) keyboard.
- Epson Model MX-100, 136 column dot matrix printer with an RS 232 interface port.

This system uses the following software:

- NorthStar BASIC, DOS, and MONITOR and with an expansion to use the NorthStar multi-user, multi-task operating system.
- Crosstalk, a software package provided by Microstuf Inc. is used for file transfers between the NorthStar microcomputer and the Amdahl mainframe computer.

Summary of applications of the above described systems.

The Amdahl V-5 mainframe is used primarily as a vehicle to access magnetic tape devices; all final archiving of Jornada LTER data is being done at this level. Our reasons include standard security and backup services; access to SAS, telecommunications, and each other (all of the above described systems communitate with this mainframe); and other services that desktop and mini-computers are unable at this time to provide. We have, however, reduced our use of this mainframe computer from 100% 18 mos ago, to about 10% at the present time. It is not clear that this is cheaper, but we now have much better control of our computing, as well as a high degree of flexibility.

The NORTHSTAR computers have proven quite effective for routine data and word processing; in addition, a locally developed software package (Ludwig) now provides for on-line bit-pad entry of plant-cover-dynamics-photo-plot data. This program allows simple outlining of plant species from projected 35mm slides, it integrates cover, and stores all relevant data for subsequent transfer to mainframe tape storage.

The H-P/1000 is a highly talented mini which allows any level of local customizing that the operators are capable of imposing. This system has a local workstation (a 2642 data and word processing 'terminal' with off-line disk storage) on which most LTER data entry and initial verification is accomplished. This device also communicates with the NMSU mainframe. This H-P/1000 computer is currently being configured into a 'data analysis and modeling workstation laboratory' under the sponsorship of the NMSU Computing Research Laboratory (where Conley holds a half-time appointment). One goal is a multi-user computing environment where researchers may conduct intensive data snooping, analysis, and presentation workshops, without the need to pause for a week while those who can do the programming try to capture the 'group opinion'. Most current programming effort on the Jornada LTER site is being expended on this system. A more specific task for this computer is the design and development of a population

modeling package that combines modules that provide knowledge-based data retention, data reduction and analysis, simulation projections and monte-carlo decision gaming, and a sophisticated user-interface. These separate modules will be driven by an AI module that will provide the services of an 'Expert System' that is capable of benefitting from erroneous predictions. Thus, this computing laboratory serves several closely related research efforts.

DATA SET HEADERS

All data sets are kept in files according to data set type and calender year. The annual breaks provide for maximum size of our files of about 5-6000 lines. Each file contains an embedded header that describes the history of the file. Every change to a file is recorded in the file header. In some cases, individual researchers manipulate their own data, and provide a reviewed and revised (if needed) file back to the DMO. In other cases, the DMO does the keying; this latter case requires a written request which is archived with the paper copy. Our data header looks like this:

```
{********************** CHANGE LOG ***************************** }
{ REVISED BY DMO latest date of revision }
        .
        .
        .
{ REVISED BY DMO earliest date of revision }
{ FILEID __-name-_______ DATA INCLUDES: }
{ STARTING WEEK NO. =              }
{ LATEST WEEK NO.  =              }
{ DATA DOCUMENTATION FORM DATED day mo yr }
{ FREQUENCY                       }
{ RESPONSIBLE P.I.
{ REVIEWED AND REVISED BY _(-PI or agreed rep only-)__ DATE day mo yr
        .
        .  ( this sequence is earliest on top, latest on bottom)
        .
{ REVIEWED AND REVISED BY ___________________ DATE day mo yr
{ SPECIAL INSTRUCTIONS:
{ DATA FIELD = ____________ INDICATES PROBLEM
{ -anything else you want to say regarding special instructions-
{ OUTSIDE DATA-ACCESS HISTORY:
{ ********************** END CHANGE LOG ***************************
```

All applications programs are being written so as to recognize the left { in col 1 as a 'comment' line. Thus we are able to include imbedded statements anywhere in the data file. A special program strips the comments from a work file for submission to canned programs that might choke on such non-data inclusions.

DATA DOCUMENTATION

A copy of our data documentation form is appended. See also the 'LTER

short form'. During 1984, an objective is to bring all LTER sites on-line with a keyworded directory, and brief description of all data sets at all sites.

CRITICAL SELF-EVALUATION

Idealisms of the past tend to be replaced by realities of the present. Not all of the Jornada LTER data sets are handled by the DMO because not all of the researchers respond to general data management criteria in the same manner. There has also been no way to assure that all data sets are routinely documented, updated, and reviewed, in a timely manner. The tendency is for people to become interested in their data when it comes time for publication and/or review, and to not think about it in the interim. This makes review times particularly hectic, with the usual concommitant danger of human error in manipulating many data sets. We have built in a redundancy system to protect against disaster in this realm; we can always return to the last step.

At the present time our data are simply stored by data type and calender year, in the form of standard ASCII files and in easy to review two-dimensional tables. They invariably contain descriptive headers that, while varying in detail, provide adequate information for linkage to researchers and documentation. A separate research project (to Conley) will be setting up selected data sets in a formal DBMS, for use in an effort at designing an 'artificially intelligent population model'. This experience will feed back to the rest of the project as a prototype (for us!) attempt to bring up such data in a manner that really works.

We are emphasizing use of canned software where feasable, and, for programs written in the DMO, professional standards of program documentation. The DMO has no method of assuring that all site programmers adhere to any given procedure.

Final testing is progressing on an on-line directory, key-worded documentation reporting program. First quarter 1984 should see the finalization of this system. A keyword searching package is being rewritten in Pascal (from FORTRAN) will provide a vaccine for the mnemonic plagueing of the current program.

File: *LATADOC*

NEW MEXICO STATE UNIVERSITY

NSF/LTER -- JORNADA SITE -- DATA SET ABSTRACT FORM

All Jornada Desert Site *NSF/LTER* data sets must be documented on the following form. For purposes of definition, this qualification applies to any data set intended for permanent archiving. It is thus assumed that <u>all</u> raw data collected under the auspices of the Jornada *NSF/LTER* program will be described by the following standardized form. In addition, supplemental data sets must also be described on this form to the maximum extent possible.

DATE: DAY_____MONTH____YEAR____(of filling out this form)

*SUGGESTED FILE I.D.:*_____(an 8 character alphameric file identification mnemonic)

*ASSIGNED FILE I.D.:*_____(leave blank: Data Management Office will coordinate assignment)

BRIEF TITLE: (unscramble the mnemonic suggested above)

INFORMATIVE ABSTRACT: Describe data set

KEYWORDS: (keywords that describe data set; maximum of 17)

*RESPONSIBLE PRINCIPAL INVESTIGATOR:*_____

RESEARCHERS: List all personnel who will be obtaining data with their initials following in "( )".

ADDITIONAL PRINCIPAL INVESTIGATOR(s):_____

RESEARCH ASSOCIATES/ASSISTANTS_____

_____

_____

SITE INFORMATION: Control___, Treatment___, Other___(describe)
(Relevant data sets used in LTER site analyses will be archived if feasable, e.g. adjacent weather data, etc.)

METHODOLOGY:

SAMPLING PROCEDURES: (Provide sufficient detail such that an aware reader could repeat the described data collection procedures).

KEY LITERATURE: (citations that describe sampling procedures)

PRELIMINARY TREATMENT OF DATA: (NOTE: It is anticipated that entered data is "Field Raw", i.e. as collected. In a few instances, such data are converted or transformed prior to recording for LTER data entry. All data conversions of this sort must be generally agreed on by Site P.I.'s, and must be adequately documented here. (Use an appendix page if required.)

ATTRIBUTES MEASURED AND FORMAT:

NOTE: The first 18 columns are proprietary identification codes for the LTER Desert Site Data Management Office. This requirement provides for collation of derived data sets by Data I.D., Time, and/or Location.

VARIABLE NAME	COLS	FORTRAN FORMAT	MEAS. UNIT	ATTRIBUTE DESCRIPTION
FILEID	1- 8	A8	N/A	FILE IDENTIFICATION CODE
	9	BLANK		
WEEKNO	10-13	I4	N/A	WEEK NUMBER (RIGHT JUSTIF
	14	BLANK		
LOCATION	15	A1	N/A	T=Treatment, C=Control
STATION	16-17	I2	N/A	STATION NUMBER
CARDNO	18	I1	N/A	CARD NUMBER 1-9
(Columns 19 - 80 your data columns)				

NOTE: If you require more than columns 19-80, use a second (etc.) card, repeat columns 1-18 above on each card.

NOTE: If your location is off the two established transects, the Jornada DMO will ASSIGN A LOCATION CODE, and will help your designate sub-location codes in columns 19 - 80.

NOTE: Somewhere in Columns 19 - 80 there must be a 1 column "TROUBLE" FLAG (0 = No Trouble, 1 = Trouble) that points to appropriate field notes describing the problem.

DATE DATA STRING COMMENCED: DAY:___MONTH:___YEAR:___.
WEEK NO._____

DATE DATA STRING TERMINATED: DAY:___MONTH:___YEAR:___.
(If ongoing, put "continuing").

FREQUENCY OF MEASUREMENT:

TIME LAG REQUIRED TO PROCESS COLLECTED SAMPLES:

EXPECTED DURATION OF STUDY:

METHODS OF RECORDING: (Field data sheets, instrumental, electronically, etc.)

COMMENTS: Include any comments here that more fully describe this data set.

The following abbreviated form can be completed from above information.

NSF/LTER SITE NETWORK - DATA SET DOCUMENTATION: NETWORK DIRECTORY

TITLE OF DATA SET:

ABSTRACT - DESCRIBE DATA SET

KEYWORDS THAT DESCRIBE DATA SET (MAXIMUM OF 16)

LTER SITE: (CIRCLE ONE) ANDREWS, NIWOT, CENTRAL PLAINS, JORNADA, CEDAR CREEK, NORTH LAKES, ILLINCIS RIVER, COWEETA, NORTH INLET, OKEFENOKEE, KONZA.

SITE COORDINATOR

NAME:_____

ADDRESS:_____

_____

PHONE: AC____/_____

PRINCIPLE INVESTIGATOR RESPONSIBLE FOR DATA SET:

NAME:_____

ADDRESS:_____

_____

PHONE: AC____/_____

SAMPLING FREQUENCY:_____

DATE DATA STRING COMMENCED

_____DAY_____MO_____YEAR

DATE DATA STRING TERMINATED

_____DAY_____MO_____YEAR
(LEAVE BLANK IF ONGOING)

REFERENCE TO COMPLETE DOCUMENTATION AT SITE: (I.E. FILE IDENTIFICATION OR DATA SET CODES)

COMMENTS:

Illinois

LARGE RIVER LTER DATA MANAGEMENT (FIELD STATIONS) UPDATE:

1) Software development and use (these run on an Apple II+)

Software	Function
DataFactory 5.0	entry, editting, listing sorting, searching, creating subsets, changing formats
DF-DC	Converts DataFactory 5.0 files to DataCapture 4.0 text files
DataCapture 4.0	Transmit text files to Champaign (CYBER or PRIME)

2) Identification and modifications of accessory data sets -

ie. Historical Illinois River Electrofishing Data.

3) Coping with change (progress?)

How do you keep data sets compatible in the face of changing goals and consequent new or modified variables?

BOB SINCLAIR ILLINOIS DEPT ENERGY & NATURAL RESOURCES
STATE WATER SURVEY DIVISION
605 E. SPRINGFIELD AVE
PO BOX 5050, STATION A
CHAMPAIGN, ILLINOIS
61820-9050

Current Configuration

Field Station	Hardware	Software	Champaign Hardware	Champaign DM software
Grafton	Apple II+ 2 drives modem (300) Silentype printer 80 col. card	DataFactory 5.0 DF-DC	IBM	
WIU-Macomb	Apple II+ w/ 1200 modem	DataCapture 4.0	CYBER	S.I.R.
	Basis (Apple compatible) 2 drives Prowriter printer		PRIME	ARC/INFO
Havana	Tektronix 4051 hardcopy unit plotter 300 B tape drive, data communications module	Self-written entry, correction and analysis programs		

Planned Configuration

Field Station	Hardware	Software	Champaign Hardware	Champaign DM software	Champaign ad. software
all	AppleIIe 2 drives Graphics terminal plotter CP/M 1200 Baud modem	DataFactory 5.0 DF-DC DataCapture 4.0	PRIME (CYBER,IBM)	ARC/INFO	SAS,SPSS, etc.

KSL 19831104

Historical Illinois River Electrofishing Data:

Researchers: Starrett, Sparks, Lubinski

Locations: 27 stations on Illinois River, but some stations missing in selected years

Period of Record: 1957 - 1970; 1973 - 1979; 1982 - 1983. (apprx. 60,000 records)

Consistency problems within data set:

a) missing stations, years
b) changing habitats over project period at same station
c) addition of new variables in later years
d) inconsistent level of detail for observations (mostly "comments")
e) early attempts at "computerization" emphasized "customized" programs, data format dependent on position of records in file.

KSL 19831104

DATA BANK MANAGEMENT REPORT

OKEFENOKEE SWAMP LTER SITE

INSTITURE OF ECOLOGY

UNIVERSITY OF GEORGIA

ATHENS, GA - 30602

Prepared by R.E. Hicks and D.H. Kearns for the LTER Data Managers meeting held at Oregon State University, Corvallis, OR - November 6-9, 1983.

PHASE I: Design and Implementation of Data Archiving Protocols and a Simple Data Management System.

This stage is nearly completed at this time. Magnetic tapes were chosen for long-term archiving and a simple management system was decided upon because of the cost effectiveness of this medium, the current lack of interactive computer equipment and software, continued compatibility with file handling and statistical software at the University of Georgia, and the identification that at present we do not need daily updates of or instantaneous access to the data base. Tasks completed during this first phase included:

1. an inventory of historical data sets.
2. the elucidation of data formats for some historical data sets.
3. a data submission policy and system, including standardized data set forms and site maps.
4. initiation of yearly interviews with Okefenokee Swamp LTER investigators and graduate students to remain abreast of ongoing data collection and familiarize investigators with the data management system. and
5. software development to streamline and standardize data archiving, management and retrevial.

Three pieces of software have been designed by the previous data manager, Dr. Randall E. Hicks. ARCHIVE is a program designed to place a 'letterhead' and directory on each archive tape. DOCPUT requires selected input from data set forms and places summarized documentation for each data set on an archive tape file immediately preceeding the

2. recovering and archiving historical data sets which are similar to LTER goals and data at the site.
3. initiation of our redundancy program to create a back-up system of data bank tapes. and
4. write protecting completed archive tapes in the data bank.

PHASE III: Interfacing with Microcomputers

A microcomputer sytem is currently being designed for use by the data management program and Okefenokee swamp investigators; then, it is envisioned that protocols will be implemented so individual investigators can access the Okefenokee Swamp Data Bank. This will allow each investigator to manage small portions of the computerized data base which are of interest to him/her more intensively and interactively using their microcomputer systems and data management software.

KSU
KANSAS STATE UNIVERSITY

Division of Biology

Ackert Hall
Manhattan, Kansas 66506
913-532-6615

31 October 1983

TO: LTER Data Managers

There will be a "Symposium on Computers and Statistics in Benthic Studies" as part of the meeting of the North American Benthological Society to be held in Raleigh, NC, during May 23-25, 1984. This will include a contributed paper session, and it occurred to me that I (or another data manager active in NABS) could present either a general paper on data management or a paper which emphasizes the progress made in data management in the LTER program. The latter appeals the most to me because it provides an opportunity for us to share with the scientific community some of the ideas and approaches which are being developed in the LTER program.

I have spoken with Todd Folsom, the organizer of the symposium, and he said such a paper would be appropriate and very welcome. He said there would also be a poster/demonstration session and encouraged me to consider a poster or microcomputer display in addition to a presented paper (this might be a good idea since contributed papers are only 12 minutes plus 3 minutes for questions).

I would appreciate any comments or ideas you may have about such a presentation--e.g., whether it is appropriate to try to represent the "LTER approach". I also welcome potential co-authors for either a paper or a poster. Perhaps one site with an aquatic-oriented research program might want to present a paper or poster highlighting its own site-specific efforts. Another idea might be to put together a general-purpose display which would be "suitable for re-use" at other society meetings as similar sessions develop in the future.

These are ideas I would like to discuss with interested people at the workshop in Corvallis. I have attached a first attempt at an abstract for a presented paper, as well as descriptive material about the symposium. Deadline for abstracts is November 30.

Marty

Marty Gurtz
Data Manager, Konza Prairie

Draft of abstract for N.A.B.S. meeting.

"Data management efforts in the Long-Term Ecological Research program"

Any data base, regardless of size, should be properly managed in order to preserve its integrity and assure its future value to researchers. Poor management procedures can too easily result in data which are either inaccessible or uninterpretable. Data management has been given a high priority in the Long-Term Ecological Research (LTER) program of the National Science Foundation. At present, 11 sites representing varied ecosystem types are investigating long-term ecological phenomena. Ideally, data management begins prior to data collection and includes design of raw data sheets, entry of data into the desired storage form, data verification, archiving, and maintenance. Because of the planned longevity of this program, it is especially important that present and future researchers, including potential collaborators from other institutions, be able to: identify existing data sets, obtain information concerning the origin of those data, and access those data (with appropriate permission). Each site, therefore, is preparing an inventory, directory, and catalog of LTER data sets. Proper documentation [personnel, experimental design, sample collection and analytical procedures, computer programs, storage formats and retrieval information, and histories of updates and usages] is an integral part of data management. While individual LTER sites vary widely in their computer (and other) systems for data storage and retrieval, these principles of data management are applied in common at all sites. Data managers from all sites have met annually to review and evaluate progress toward quality site-specific data management and to continue discussions on the convergence of systems of coordinated database management for intersite exchange.

CALL FOR PAPERS

NABS SYMPOSIUM ON COMPUTERS AND STATISTICS IN BENTHIC RESEARCH

A methodology symposium is being planned by Ron Hellenthal and Todd Folsom for the 1984 meeting where participants can show how they have handled problems in data analysis and management. The goals of the symposium are to show 1) how new computer hardware and software can improve benthic data analysis and interpretation, and 2) how various statistical methods performed by computers can extract information from data. The symposium will be a concurrent session. There is the possibility of a poster/demonstration session where microcomputers can be set up or detailed output displayed. We also expect to have some computer companies in the Raleigh area as exhibitors.

Persons wishing to contribute to this symposium should prepare a detailed abstract and submit it in the usual way. A copy should be sent to: Todd C. Folsom, Duke Power Company, Production Environmental Services, Route 4, Box 531, Huntersville, NC 28078.

TODD C. FOLSOM, Ph.D.
Aquatic Ecologist

DUKE POWER COMPANY
ENVIRONMENTAL LABORATORIES
Rt. 4, Box 531
Huntersville, N.C. 28078
(704) 875-1381

1. Introduction: W. Michener

2. Optional introductory paper: Paul Risser

3. IBP data management perspective; successes and failures:
   - Fred Wagner (logan, Utah) c. W.C.
   - Freeman Smith c. W.C.
   - Dick Waring c. S.S.
   - Paul Risser

4. Rdm protocols in LTER: Paul Alabach

5. Data entry and quality assurance: c. Susan Stafford

6. Data entry and quality assurance: NOAA, ORNL, EPA

7. Cataloging and documentation:
   - Jornada
   - Andrews
   - N. Lakes
   - K.B.S.

8. " " "

9. Data retrieval and security

10. Data communication; problems and promise in inter-computer communication: Mark Klopsch

11. RDM; large vs. small: Andrews, epa, ORNL vs. K.B.S., Konza, Ill., Jornada

12. Example of use of archived data, data synthesis, problems: N. Lakes, K.B.S.

13. RDM; Historic data sets: Carl Bowser

14. Data inventories; regional, national: Allen, O'Neil c. Mel Dyer

15. RDM; a consultant's perspective (output oriented): Rod Strand c. W.M.

16. Graphics, cinematography: ORNL c. M. Farrell

17. Ecological modelling: George Innis c. W. Conley

18. Ecological modelling: Shugart c. Mel Dyer

19. Modelling; theoretical data integration: c. Mel Dyer
   - Allen (Univ. Wisconsin)
   - O'Neil (ORNL)
   - Sugihara (ORNL)

20. Professional Workstation: Walt Conley

21. Scientific data base characteristics, comparisons: Vera Komarkova

22. Networking: Gary White (Los Alomos) c. W. Conley

23. New technology for data acquisition: Kathleen Zinnel

24. Small lab. automation: M. Marozas

25. Meteorolgical data (Data logger) NOAA

26. Photography/digitization: John Ludwig

27. Satellite remote sensing (Landsat, for example):
    Charles Showen (USGS) c. Bob Sinclair
    James Tucker (Goddard) c. Mel Dyer

28. Geographic information systems: Bob Sinclair

29. Critique of rdm:
    - Walt Conley
    - Mike Farrell

30. Future directions:
    - Mel Dyer
    - Dick Marzolf

TOPICS TO BE CONSIDERED : FALL / WINTER '84

HOBCAW GEORGETOWN, S.C.

I. BACKGROUND

A. DATA

1. definitions
2. types - (structure, etc.)

B. RDM

1. historical perspective
2. range of current applications
3. future directions
4. lay framework for following contributions

E. RDM CONCEPTS & METHODOLOGY (inc. hardware / software considerations)

A. Experimental design
B. RDM STRATEGY
C. DATA ENTRY
D. QUALITY ASSURANCE
E. CATALOGING / DOCUMENTATION
F. Security
G. DATA RETRIEVAL
   1. STATS
   2. GRAPHICS, CINEMATOGRAPHY
   3. MODELLING
H. LONG TERM DATA STORAGE

III. DATA EXCHANGE

A. DATA SET INVENTORIES (up to regional & national level)

B. MECHANISMS

1. integration of data sets (on several levels)
2. documentation
3. networking of sites

V. APPLICATIONS

A. RDM in relation to remote sensing data.

B. Laboratory automation

C. Large scale data manipulation (special considerations)

D. Sample tracking

E. OUTPUT-ORIENTED RDM (i.e. the consultant's environment

F. DATA SECURITY (new applications)

G. Q.A.

H. RDM with extant data sets

I. FUTURE OF RDM

A. GOALS

B. DIRECTIONS

Andrews

11 October, 1983

Forest Science Data Bank Data Cataloging System

A. Introduction

The data catalog for the FSDB forms the central framework for documenting and organizing the the flow of data and its associated documentation in the department. The data catalog system as developed in 1981 has four key functions:

1. to serve as the primary location of the most current information on the location and organization of all of the data sets and datafiles maintained by the FSDB.

2. to produce an easy-to-read printed catalog of data sets and associated datafiles maintained by the data bank to aid data retrieval for all potential users of the FSDB.

3. to produce an annual report to principal investigators contributing data to the FSDB; including information on which of their data we have, and what documentation is available.

4. to allow for automatic updating of the FSDB data retrieval system on CYBER. This is accomplished by generating a compact-format file containing the names and locations of all tapefiles in the FSDB.

Demos

DATASTAR city Data Base

STATPRO

INGRES

Up until October, 1983 this system was totally dependent on in-house software developed by Paul Alaback. Programs were written in FORTRANIV, FORTRANV, CBASIC-80, and Pascal MT+ to implement this system. These programs served a useful purpose in providing a functional relational data base management and data bookkeeping system when no such software was commercially available to the FSDB. With the installation of the INFOSTAR relational database management system developed by MicroPro International, Inc in September 1983, the FSDB was presented with the opportunity of incorporating all of the needed features of the data cataloging system within the framework of one simple system.

The principle advantages of the new system are as follows:

1. The entire system works within the INFOSTAR package so that all features are documented in detail in the reference manuals.

2. The system is simple to use. The component programs (WORDSTAR, DATASTAR, FORMGEN, REPORT, REDIT, FORMSORT, and SUPERSORT) are completely intercompatable having the same commands and similar philosophy of operation so that once a user has mastered the operation of one program it is simple

to learn the operation of the other programs.

3. All datafiles are stored in a random-access format so that data editing and retrieval is fast and easy (it takes the same amount of time to retreive any individual record in the database)

4. With little or no modification the system can be used as a 'on-line' data directory, whereby potential users of the data bank can easily retrieve up-to-date information on the status of any data set maintained by the data bank by simply entering a few commands.

In the following sections the procedures needed to maintain the cataloging system are described. For more details on how to use the DATASTAR program consult the DATASTAR reference manual chapters 3. (Data Entry) and 4. (Data Retrieval and Modifications).

B. How to add new data set codes to the catalog system

1. Put the 'PROGRAMS' disk in drive A:. Put the data directory disk in drive B:. Enter **^C** (control and C) to tell the computer you have changed disks; and **B:** (return) to change the default disk to the B: drive. Enter **A:DATASTAR TITLES** (return) to invoke the system.

Note:'TITLES' is the name of the series of files that contain information on dataset titles and data set documentation (TITLES.DTA=raw data, TITLES.NDX=key field index file, TITLES.DEF=data entry form overlay file, see DATASTAR manual).

NONE OF THESE FILES SHOULD BE EDITED DIRECTLY. ALWAYS MODIFY THESE FILES USING THE DATASTAR SYSTEM (otherwise a whole database can be scrambled; requiring that FSDB personnel resort and recompile the whole series to make the files useable again).

2. Determine what number the new dataset code should have:
a) enter **K** to get into the 'key' mode and the name of the first dataset code in the next dataset category (see index page of the printed data directory) (and a **return**). The record should appear on the screen.
b) retrieve the last data set code of the previous data set category. Enter **^EI**. This command exits the current mode (^E) then puts you into the scan by key index mode (I). Enter **^P** to view the previous record. The record displayed should be last number in the data set category. To make sure you are looking at the correct record enter **^N** to view the next record. This record should be the first number of the next data set category. The new data set should be given the next consecutive number in the data set category.

Example: a new data set code is needed for the TV series. Enter

ZD001 (the first code in the next data set category)
^EI
^P

TV005 should appear on the screen. The next data set code in the TV series should be TV006

3. switch modes from the index mode to the 'add' mode by entering **^EA.**

4. Enter the required information into the data form. If you have entered all of the needed information, enter **^B** to skip to the end of the form and a **return** to file the data. Note: if you have a non-standard data set code (one in which the first two letters of the data set code are not equal to the name of the data set category) the program will not file the new data until you enter the correct data set category code.

5. If new tapefiles are to be added enter **^EE** then **DATAFILE** in response to the question 'enter name of form'. Hit the space bar to enter the add mode.

6. When done adding new records enter **^EEC** to exit the system.

C. How to add new tapefiles to the data directory system

1. Enter **A:DATASTAR DATAFILES** (return) and hit the space bar to invoke the system and enter the datafile database in the add mode (unless you are already in the system as in A part 5 above).

Note: all of the files containing information on the names, locations, and documentation of the individual datafiles associated with each dataset in the FSDB are given the name DATAFILE.xxx. **DATAFILE.DTA**= the raw datafile, **DATAFILE.DEF**=the data entry form, **DATAFILE.NDX**=the key field index file. NONE OF THESE FILES SHOULD BE EDITED DIRECTLY. ALWAYS MODIFY THESE FILES USING THE DATASTAR SYSTEM (otherwise a whole database can be scrambled); requiring that FSDB personnel resort and recompile the whole series of files to make them useable again).

2. Enter the required information onto the form. Note the tape VSN number and security code are on the bottom of the form. To get to these fields quickly enter **^L** (this will take you to the last field in the form). When done hit **^B** to get to the end of the form (if you are not already there) and a carriage return to file the new data. Note: if you have a non-standard data set code (one in which the first two letters of the data set code are not equal to the name of the data set category) the program will not file the new data until you enter the correct data set category code.

D. How to retrieve and/or update information that has already been entered

1. Enter the database that you need to modify (**A:DATASTAR TITLES** or **A:DATASTAR DATAFILE**).

2. Enter the key mode (enter a **K**).

3. Enter the name of the data set code to be modified and the study id (if present) and **return**. The record to be modified will appear on the screen. Use the cursor positioning commands to move to the field(s) that need to be modified. Write over the obsolete information. You only need to change those characters that are wrong (not the whole field). When done making changes enter ^B and a **return** to file the updated record.

4. When done editing enter **^EEC** to exit the system.

AD 1 Aquatic: substrate decomposition
AN 2 Aquatic: nutrients (see also Central chemistry laboratory)
AS 3 Aquatic: secondary production and biomass
CL 4 Central analytical chemistry laboratory: miscellaneous
IA 5 Central analytical chemistry laboratory: stream samples
IT 6 Central analytical chemistry laboratory: soil solution, throughfall
IS 7 Central analytical chemistry laboratory: soil cores
IE 8 Geology: erosion chemistry samples
IB 9 Geology: erosion mass samples
GS10 Geology: stream Geomorphology
GV11 Geology: vegetation-erosion relationships
HP12 Hydrology: precipitation (see also Meteorology)
HS13 Hydrology: snowfall (see also Meteorology)
HM14 Hydrology: soil moisture and conductivity
HF15 Hydrology: streamflow data
MG16 Meteorology: miscellaneous
MH17 Meteorology: hourly summary
MD18 Meteorology: daily summary
MM19 Meteorology: weekly or monthly summary
TD20 Terrestrial: organic decomposition
TL21 Terrestrial: litterfall mass
TN22 Terrestrial: nutrients (see also Central chemistry laboratory)
TP23 Terrestrial: vegetation(see also silviculture)
TS24 Terrestrial: fauna
TW25 Terrestrial: plant-water relationships
SA26 Systematics: arthropods
SB27 Systematics: birds
SC28 Systematics: bryophytes
SF29 Systematics: fish
SI30 Systematics: fungi
SM31 Systematics: mammals
SV32 Systematics: vascular plants
BZ33 A. Berg thinning studies: site description and experimental design
BE34 A. Berg thinning studies: tree diameter talley,height and thinning
BD35 A. Berg thinning studies: tree volume (dendrometer)
BC36 A. Berg thinning studies: tree mapping data
BH37 A. Berg thinning studies: felled tree data
BU38 A. Berg thinning studies: seedfall and thinning litter biomass
BL39 A. Berg thinning studies: understory vegetation studies
BP40 A. Berg thinning studies: precipitation data
BT41 A. Berg thinning studies: duff temperatures
BP42 A. Berg thinning studies: fuel moisture
BS43 A. Berg thinning studies: soil moisture and temperature
RA44 Forestry Intensified Research in SW Oregon: adaptive data
RF45 Forestry Intensified Research in SW Oregon: fundamental data
FE46 Forest engineering studies
FG47 Forest genetics
FP48 Forest protection
FS49 Silviculture
CR50 Cooperative Research on Alternative Forestry Treatment Systems:(Walstad)
CS51 Cooperative Research on Alternative Forestry Treatment Systems:(Newton)
ZM52 Miscellaneous: other study data
ZD53 Miscellaneous: data bank administrative and organizational files

Andrews

A DEMONSTRATION OF THE USE OF THE ON-LINE FOREST SCIENCE DATA BANK DATA DIRECTORY AND DATA CATALOG

I. Introduction

Purpose--The Forest Science Data Bank maintains over 380 titles of data from studies conducted by faculty in the Department or by joint inter-institutional projects conducted throughout the Pacific Northwest and Alaska. These data have been archived and the documentation on both the research origin and the data formatting details have been organized to facilitate their use not only by faculty who initiated these studies, but by other scientists that could use the information for new types of analysis, to formulate new studies, or for comparison with new data. Oftentimes researchers take preliminary (or sometimes extensive) data on a topic only to find that some archived data is highly relevant to their work which if expanded upon could provide a superior basis for testing the hypotheses of interest than a new, more limited study. Increased utilization of extant data has great potential for increasing research efficiency and providing the means to make more robust tests of hypotheses. The National Science Foundation, and similar organizations have consistently pointed to these reasons to justify the funding of data management systems at leading research institutions across the country.

The prime impetus for the development of the Forest Science data bank cataloging system has been to make it easier for uninitiated users to 1) identify which data sets may be relevant to their needs or interests; 2) to easily determine what the status is of the studies in question (i.e. what level of documentation is available, how many datafiles are associated with the study, how are the datafiles organized, when was data last added to the data bank); and 3) provide a bookkeeping mechanism for data that are being entered and processed or stored by the data bank and the quantitative services group.

Catalog organization--The on-line catalog is organized into three main databases: 1) data set titles and documentation status (the **TITLES** database) 2) datafile locations and descriptions (the **DATAFILE** database) and 3) data formatting and data coding information (the **FORMATS** database). The three databases are tied to one another by the 'dataset code'. The dataset code is composed of four to seven digits that uniquely identify each dataset in the data bank*. After determining the dataset code(s) of interest from the master database, (**TITLES**) more detailed information on datafiles that are associated with the dataset and names of variables and their format can be obtained from the other files** .

---

* Study code is a code that uniquely identifies each of a series of studies that are within one dataset code. Most datasets do not have study id's since their is only one study within that dataset code.

** The FORMATS file has just been implemented. At the present time, most of the data formatting information is located in paper file folders in FSL 204 or 248 or is on datafile headers. We plan to transfer this information to the FORMATS file as soon as possible.

II. How to use the system

A. Set up computer.

Put floppy disk labeled "Forest Science Data Bank Catalog" in disk drive "A". Enter control and C together(¢C).

B. Invoke the DATASTAR database management system.

Enter:

DATASTAR (name of database) and carriage return

To search for a dataset code, name of database = TITLES. To retrieve information on location and organization of datafiles, name of the database = DATAFILE. To look at data formatting and coding information, the name of database = FORMATS.

C. In response to the question "Mode desired" enter a **K** to go into the search by key mode. Enter the name of a dataset code (for example TV003) and two carriage returns or the dataset code (and carriage return) and the study id (if present) and another carriage return. The information on that dataset should now appear on the screen.

To look at another record enter **control** and **B** and a **carriage return.** Then enter the name of the next data set code (and study id if present). If you wish to look at series of datasets (i.e. TV001, TV002, ...) it may be more convenient to enter the **I** (scan in index order) or **D** (scan in data file order) modes. Enter **control** and **EI** to enter the index mode or **control** and **ED** to enter the data file order scan. Once in either of these modes the next record in the file can be viewed by entering a **control** and **N.** The previous record in the file can then be seen by entering a **control** and **P.** To return to the key mode enter **control** and **EK.**

For more information on how the DATASTAR database management system works consult the DATASTAR REFERENCE MANUAL, chapters 3 and 4 (mostly 4).

III. A demonstration

a. How to retrieve information from the data titles database:

1. Put "Forest Science Data Bank Catalog" disk in drive A.

2. Enter control and "C" (¢C)

3. Enter **DATASTAR TITLES** (carriage return)

4. Enter **K** to enter the search by key field mode.

5. Enter **TP56** (carriage return) and **LTV** (carriage return) to retrieve information on dataset code TP56, study category LTV. Documentation on the long-term ecological research project should appear on the screen.

b. How to view a user defined subset of data titles: the use of the edit scan mask.

1. To see all dataset codes with J.F. Franklin as the principle investigator, enter the edit scan mask mode (**¢EM**). Hit the carriage return until you reach the principle investigator field. Then enter **Franklin,J.F. ¢B** and carriage return. Note: this scan mask is in effect for all the other modes in DATASTAR until you go back to the edit scan mask mode and change it.

2. Enter the scan by key field index mode. Enter **control** and **EI**.

3. To view the next record in the file (alphabetized by data set code and study id) enter **control** and **N**, to view the previous record enter **control** and **P**. Only those datasets associated with J.F. Franklin should appear on the screen.

c. How to switch forms to get more detailed information

1. Enter a **control** and **EE** and **DATAFILE** (carriage return) to exit the current database (TITLES) and enter a new database (DATAFILE). Note: the first control E exits you from the current mode, the second exits you from the current form, or database.

2. Enter a **K** to get into the search by key field mode.

3. Enter **TP56** (carriage return) and **LTV** (carriage return) to retrieve information on dataset TP56 study id LTV.

4. Enter a **control** and **EI** to exit the key mode and enter the scan in key index order mode.

5. By entering **control** and **P** or **N** documentation on the next or the previous datafile associated with TP56 LTV will be displayed on the screen. Note: if you scan beyond the number of datafiles associated with TP56 LTV in either direction you will start seeing datafiles associated with other dataset codes (i.e. TP56 DRYD or TP56 REFS).

6. Notice that in one of the fields, the numbers or codes of each data format type (card type) contained in each datafile is listed. For the datafile LTVSITE for example, the only data format type present is SITE.

7. To see what variables are contained within the datafile LTVSOIL exit the DATAFILE database and enter the FORMATS database: enter **control** and **EE**, and **FORMATS.**

8. **Enter** the edit scan mask mode by entering **M.**

9. Enter **TP56** (carriage return) **LTV** (carriage return) in the dataset and study id fields, and **SITE** in the data format type field. Only those variables associated with dataset TP56 LTV and format type SITE will be displayed on the screen. Enter a **control B** and carriage return to exit the edit scan mask mode.

10. Documentation on the first variable in the format type should appear on the screen. To see the other variables in this format type, switch to the scan by key index order mode (**control EI**) and enter **control N** (each time you enter this code the next variable in the file will appear on the screen until you reach the end of the file).

d. To exit the system:

Enter **control EE¢C** or turn the computer off.

Note: all commands mentioned in this demonstration are listed on the "menu" located at the top of the terminal screen by the DATASTAR system to assist new users in determining codes functions. See DATASTAR reference manual chapters 3 and 4.

FORMATS FORM LISTING AND FIELD ATTRIBUTE DEFINITIONS

FIELD NUMBERS

831020

FOREST SCIENCE DATA BANK DOCUMENTATION

DECK ID: _____1 STUDY ID: _____2 INITIALS: __3
FORMAT TYPE CODE: _____4 DATE OF LAST UPDATE: __5/__6/__7
FORMAT TYPE TITLE: MM DD YY

_____8
VARIABLE NAME: _____9 (VARIABLE EXTENSION SHOULD BE PRECEDED BY A ':')
VARIABLE LABEL: _____10
VARIABLE CODED: * (Y=VARIABLE IS CODED,' '=NOT CODED)
BEGINNING COLUMN: __12
FORTRAN FORMAT: 13 X (LENGTH OF SKIP FIELD)
* (FORTRAN VARIABLE TYPE: F=REAL,I=INTEGER,A=ALPHA)
15 (VARIABLE FIELD WIDTH)
.16 (NUMBER OF COLUMNS TO THE RIGHT OF DECIMEL,REALS ONLY)
MEASURMENT UNITS: __17
DATA TYPE: * (* = FREEFORM, F = FIXED FORMAT)
SEPERATOR: (ONLY FOR FREEFORM FILES) * STRING DELIMITER (' OR ") *
MISSING DATA CODE: _____21
VARIABLE TYPE: * (V=VARIABLE,K=KEY VARIABLE,C=CONSTANT)
MINIMUM VALUE: _____23
MAXIMUM VALUE: _____24
PRECISION: _____25
VARIABLE DEFINITION: _____26
_____27
_____28
_____29
(FOR CODED VARIABLES ONLY):

NUMBER OF CODES: 30

CODE DEFINITION

1) _____31: _____32
_____33
2) _____34: _____35
_____36
3) _____37: _____38
_____39
4) _____40: _____41
_____42
5) _____43: _____44
_____45
6) _____46: _____47
_____48
7) _____49: _____50
_____51
8) _____52: _____53
_____54
9) _____55: _____56
_____57
10) _____58: _____59
_____60
11) _____61: _____62
_____63
12) _____64: _____65
_____66
13) _____67: _____68
_____69
14) _____70: _____71

ZD001T FORM LISTING AND FIELD ATTRIBUTE DEFINITIONS

FIELD NUMBERS

FOREST SCIENCE DATA BANK DATA CATALOG SYSTEM
CARD TYPE 1: DATA SET TITLES AND DOCUMENTATION STATUS

DATA SET CODE _____1 STUDY ID _____2 PRINCIPLE INVESTIGATOR _____3
STUDY TITLE _____4
DATE OF LAST UPDATE: _5 _6 _7 (DAY MONTH YEAR)
STUDY CATEGORY _8
DOCUMENTATION STATUS:
ABSTRACT 9 (1=ON-LINE, 2=COMPLETE, 3=IN REVIEW, 4=PI NOTIFIED)
DATA STRUCTURE * (1=ON HEADER OR INFOSTAR, 2=COMPLETE, 3=INCOMPLETE)
STUDY COMPLETION * (I=INCOMPLETE, OR ONGOING; C=COMPLETE)
PROJECT CODE: * (L= LTER PROGRAM)
OTHER DATA SET COMMENTS:
_____13
_____14
_____15
_____16

ZD001D FORM LISTING AND FIELD ATTRIBUTE DEFINITIONS

FIELD NUMBERS

ZD001: FOREST SCIENCE DATA BANK DATA CATALOG SYSTEM
CARD TYPE 2: DATA FILE LOCATION AND DESCRIPTION

DATA SET CODE _____1 STUDY ID _____2 DATAFILE NAME _____3
DATA TITLE: _____4
DATE OF LAST ARCHIVAL _5 _6 _7 (DAY MONTH YEAR) CONTACT: _____8
CARD TYPE(S) _____9 _____10 _____11 _____12 _____13 _____14 _____15
_____16 _____17 _____18 _____19 _____20 _____21 _____22 _____23 _____24
STRUCTURE (FIXED OR INDETERMINANT SEQ) * NUMBER OF CHARACTERS OF DATA _____26
NUMBER OF LOGICAL RECORDS _____27
DATA COMPLETION STATUS: * (I=ACTIVE FILE, INCOMPLETE; C=STATIC FILE, COMPLETE)
DATA STORAGE: * (1=TAPEFILE W/HEADER, 2=TAPEFILE, 3=CARD FILE, 4=NON-FSDB,
7=TAPEFILE, BUT NOT YET IN DATABANK)
QUALITY ASSURANCE: * (U=UNVERIFIED, E=EDITED VISUALLY, OR BY 2-ENTRY
VERIFICATION, L=CHECKED WITH DATA LIMITS AND VERIFIED
DATA SOURCE: * (R=RAW DATA, L=LITERATURE, D=DERIVED, T=TRANSFORMED, CONDENSED)
(IF L,D,C OR T GIVE SOURCE) _____32
_____33
DATA FILE COMMENTS (INCLUDE MORE INFORMATION ON THE PARTICULAR CHARACTERISTICS
OF THIS FILE): _____34
_____35
_____36
_____37
_____38
_____39
_____40
_____41
_____42
_____43
_____44
_____45
_____46
_____47
_____48
_____49
_____50
_____51
_____52
_____53
********** FOR FSDB PERSONNEL USE ONLY **********
***** TAPE VSN NUMBER ____54 SECURITY CODE FOR RETRIEVAL ____55 **********
****

FOREST SCIENCE DATA BANK DATA SET FORMAT SPECIFICATIONS

DATA SET CODE: TP56 LTV FORMAT TYPE: SITE DATA TYPE: FIXED FORMAT

VARIABLE NAME	VARIABLE LABEL	FORMAT	COLUMNS	MISSING CODED
DATACODE	DATA SET ID CODE	A5	1 -5	
CORS2-3	% OF REMAINER GREATER THAN 2MM, H3	I2	106-107	
COLOR3	COLOR FROM MUNSELL COLOR CODE H3	A6	108-113	
PLOT	PLOT NUMBER	I4	11 -14	
CORS2-4	% OF REMAINDER GREATER THAN 2MM, H4	I2	112-113	
HORTH4	THICKNESS OF THE FOURTH HORIZON	I3	114-116	
CORS2+4	% PARTICLES GREATER THAN 2CM IN H4	I2	119-120	
COLOR4	MUNSELL COLOR CODE, H4	A6	123-128	
SEQ	SEQUENCE NUMBER	I2	15 -16	
AREAID	AREA ID	A4	17 -20	
DATE	DATE (YYMMDD)	I6	21 -26	
PLOTTYPE	PLOT TYPE	A2	27 -28	
PLOTAREA	PLOT AREA	I5	29 -33	
ELEV	ELEVATION IN METERS	I4	34 -37	
SLOPE	PERCENT SLOPE	I4	38 -41	
ASPECT	ASPECT IN DEGREES	I3	42 -44	
PRIMSLPP	PRIMARY RIDGE POSITION	A1	45 -45	Y
LINENUM	LINE NUMBER (CARD TYPE)	A1	6 -6	
DEPTHBED	DEPTH TO BEDROCK	I3	61 -63	
PITDEPTH	DEPTH OF SOIL PIT	I3	64 -66	
LITTERTH	LITTER THICKNESS	I3	67 -69	
HORTH1	THICKNESS OF HORIZON 1	I3	69 -71	
STUDYID	STUDY IDENTIFICATION	A4	7 -10	
CORS2+1	PARTICLES GREATER THAN 2CM IN H1	I2	74 -75	
COLOR1	MUNSELL COLOR CODE HORIZON 1	A6	78 -83	
HORTH2	SECOND HORIZON THICKNESS	I3	84 -86	
CORS+2	PARTICLES GREATER THAN 2CM IN H2	I2	89 -90	
CORS2-2	% OF REMAINDER GREATER THAN 2MM H2	I2	91 -92	
COLOR2	COLOR ACCORDING TO MUNSELL	A6	98 -103	
HORTH3	HORIZON THICKNESS 3	I3	99 -101	
HYDRO	HYDROLOGICAL FEATURES	A2	47 -48	5
HORTOPOG	HORIZONTAL TOPOGRAPHY	A1	49 -49	Y
VERTOPOG	VERTICAL TOPOGRAPHY	A1	50 -50	Y
SLOPEPOS	SLOPE POSITION	F4.1	57 -60	
TEXT1	TEXTURE OF FIRST HORIZON	I2	72 -73	Y
CORS2-1	% OR REMAINDER GREATER THAN 2MM, H1	I2	76 -77	
TEXT2	TEXTURE OF SECOND HORIZON	I2	87 -88	Y
CORS2+3	% PARTICLES GREATER THAN 2CM IN H3	I2	104-105	
TEXT3	TEXTURE OF THIRD HORIZON	I2	102-103	Y
TEXT4	TEXTURE OF FOURTH HORIZON	I2	117-118	Y

Don Henshaw

Introductions

11/7/83

MINN. Data base units
SPSS units

OKE Sewlace Access (ie, restricted, unrestricted)
keywords
Dates, location, etc.
Description

SAS program for means and ranges.
investigator interviews → micros in future

Illinois Spread sheet of status of data sets.
Bob Shows documentation, archiving.
Ken

"National handbook of recommended methods for water-data acquis."
NAWDEX: USGS data search

JORNADA Quality Insurance - later changes in data - ie., mistakes.
N.M. on going clean-up.
Walt. when ARCHIVED - before or after clean-up.

Embedded comments in data files.
work station with instant graphics capabilities.

NIWOT RIDGE Digitized maps.
VERA Campbell recorders.
Jim Halfpenny Data reports

KONZA PRAIRIE No historic data base
DICK, MARTY inventory / Data set catalog / Data entry prog / Discussion w/ entry person Don investigator / Archival / Analysis SAS customized prog for routine tasks / Reevaluation

BARUCH MARINE LAB HP 7221-T Plotter (supports SAS graphs)
Marv. DEC mini

11/7/83

Northern Lakes : SIR Data Base. (main UNIVAC - micros APPLE)
TROUT LAKE modelling vs. STAT ANAL.
Carl ↳ Diff eq.
Data entry prog – in house & commercial (STATPRO)
↑ PASCAL ↗

STATS : MINITAB, STATPRO, SPSS

Data automation systems – for chem lab. ISAAC system

OKEFENOKEE associated with Dan Kearns.
Bud IBM-PC ⟷ Plasma spectrophotometer
DBASE II

PAWNEE [redacted] VAX - UNIX system
Van PORTABILITY OF MAG TAPES.

Kellogg (MSU) VAX 11/70 VMS opn. system (not UNIX)
John → direction: help people do their own work.
INGRES - documentation - Data Base System.
PSTAT, BMDP
INGRES calls CR21 Data logger.

Long Term Data: maintain that which makes it long term
make changes additive.
parallel methods when change in technology occurs.

11/8/83

communication packages
'CROSSTALK' - comm packages now embedded in micros
TIMENET

CORONA - similar to IBM

SAS FILES, SAS files, header files, SIR
Accessibility
Audit-trail, change log.

'CARROTS' for PI's
1- OUTPUT - EXploratory analysis
2 - (work station concept $30,000)
3 - DM takes care of data.
4 - 'FLASH' - slick output.
Balance of statistical anal. and graphical anal.

Modelling
1 - most rigorous use of data base
2 - requires more communication between files
synthesis of variety of data.
3 - storage of 2nd order results.
4 - LISTS of hard data
5 - model aid programs
PRE-MOD, MOD-AID

* Communication of hardware, software, pubs, etc.
* DM need support of PI
* ACCESSIBILITY of prime importance
* DATA MNGMT independent of personnel
but also, short-lived software DBM systems.
* keep researchers close to data.
* DM evolving into consultants with PI's doing more data analyses in future.

US DEPT INT
USGS - WRD
Office of Water Data Coordination
Porter Ward
RESTON, VA

CATALOG / DOCUMENTATION

Qual. ins. - double entry, visual checks.

Directory - listing of data sets, how to access.

Data archiving: mag tapes.

· Asynchronous communication - block checking needed.

GOALS / NEXT YEAR

- WORKSHOP FALL '84 HOBCAW
- ABSTRACTS per form - for data sets
- Site progress reports.
- indexing non-machine readable data.
- Quality assurance - (beyond transcription)