Development and use of Soilgrasp, the Montana soil pedon database management system
by Zachary Reed Sims

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Soil Science
Montana State University

Abstract:
Soil characterization data for 838 sites in Montana existed oh magnetic tape but were not organized into a usable computerized system. These data (from 31 files) were merged with two smaller data sets into a single database file structure with 1189 sites.

GRASP (Generalized Rapid Access Software package) was chosen to manage the database. GRASP is capable of all necessary data manipulation procedures. It also performs most common statistical operations, and can pass data to other software.

A user interface program called SOILGRASP was developed to provide on-line help for researchers, guiding them through database studies. It eliminates much of the complex syntax required by GRASP by converting user responses to SOILGRASP's questions into the proper commands.

The Soilgrasp system consists of the data files, their management software, the user interface and the researchers who access the data. A supplementary user manual was written for soil scientists using the System.

The System was used successfully in a classroom situation. Graduate students conducted original research with the database for class projects.

Two scientific studies were completed with the database system. The first was an investigation of the influence parent material has on soil characteristics in Montana. Textural inheritance was evident in soils developed from igneous rock, but not in-limestone-derived soils. Sand, silt, clay and coarse fragment contents of 535 soil horizons over these rock types were summarized. Effervescence and CaC03 content of limestone-derived soils were consistent with expectations. The second study concerned site and soil properties related to organic carbon content. Organic carbon content of 130 Montana A-horizons was not significantly correlated with clay content, in contrast to soils of the Southern Great Plains, where the correlation was high. In Montana, "organic carbon content is more closely related to elevation and precipitation.

DEVELOPMENT AND USE OF SOILGRASP, THE MONTANA SOIL PEDON DATABASE MANAGEMENT SYSTEM

by

Zachary Reed Sims

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Soil Science

MONTANA STATE UNIVERSITY
Bozeman, Montana

March 1986

APPROVAL

of a thesis submitted by

Zachary Reed Sims

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

3-7-86 — Gerald A. Nielsen
Date — Chairperson, Graduate Committee

Approved for the Major Department

Mar. 7, 1986 — Dwane G. Miller
Date — Head, Major Department

Approved for the College of Graduate Studies

March 11, 1986 — Henry L. Parsons
Date — Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in his absence, by the Director of Libraries when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature Z Reed Sims

Date 3-7-86

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation for the help provided by my major professor, Dr. Gerald A. Nielsen, whose unbounded enthusiasm and support were offered at every turn in the course of this project.

Special thanks are due also to Drs. Cliff Montagne and Paul Kresge, who supplied many ideas and a certain practicality as members of my thesis committee.

Thanks also to Stephen J. Harvey, whose assistance during the beginning of the project helped me deal with that rogue, the CP-6 mainframe computer. Two others who made life with GRASP easy were Diane Antonson and Lee Faulkner, of MSU's Center for Data Systems and Analysis.

Many faculty members and graduate colleagues have contributed directly and indirectly to the successful completion of this project. These include Dr. Earl O. Skogley, Jae Eui Yang, Mark Tomer, Michael Brown, Daniel Long, Michael Larson, Mmd. A. Matin, Robert Snyder, Richard King, Carol Bittinger, Dal Burkhalter and David Harvey of the MSU Computing Services Center, Julie Fate, Rowdy Bristol, Jim Cummings, and David Webb.

My wife, Chris, deserves special thanks for her unfailing editorial, culinary, emotional, financial, and Christian support throughout the progress of this thesis.

TABLE OF CONTENTS

TABLE OF CONTENTS-Continued

LIST OF TABLES

ABSTRACT

Soil characterization data for 838 sites in Montana existed on magnetic tape but were not organized into a usable computerized system. These data (from 31 files) were merged with two smaller data sets into a single database file structure with 1189 sites.

GRASP (Generalized Rapid Access Software Package) was chosen to manage the database. GRASP is capable of all necessary data manipulation procedures. It also performs most common statistical operations, and can pass data to other software.

A user interface program called SOILGRASP was developed to provide on-line help for researchers, guiding them through database studies. It eliminates much of the complex syntax required by GRASP by converting user responses to SOILGRASP's questions into the proper commands.

The Soilgrasp System consists of the data files, their management software, the user interface and the researchers who access the data. A supplementary user manual was written for soil scientists using the System.

The System was used successfully in a classroom situation. Graduate students conducted original research with the database for class projects.

Two scientific studies were completed with the database system. The first was an investigation of the influence parent material has on soil characteristics in Montana. Textural inheritance was evident in soils developed from igneous rock, but not in limestone-derived soils. Sand, silt, clay and coarse fragment contents of 535 soil horizons over these rock types were summarized. Effervescence and $CaCO_3$ content of limestone-derived soils were consistent with expectations. The second study concerned site and soil properties related to organic carbon content. Organic carbon content of 130 Montana A-horizons was not significantly correlated with clay content, in contrast to soils of the Southern Great Plains, where the correlation was high. In Montana, organic carbon content is more closely related to elevation and precipitation.

INTRODUCTION

The State of Montana has, since 1960, acquired detailed descriptions of soil pedons from various sources, including Soil Surveys, graduate student projects, SCS Pedon Descriptions including laboratory characterizations performed by the Lincoln, Nebraska and Mandan, N. Dakota Labs, projects in cooperation with the Bureau of Land Management, and ongoing Montana Agricultural Experiment Station research.

The data are useful for research and county extension work, but were not readily accessible to soil scientists. The volume of data presented serious problems for manual access, which required a worker to page through six or more thick volumes of paper copy. Attempts to perform soils studies were limited to small sample sizes and required extensive data sorting by hand. Broad studies were difficult and seldom complete.

Manual data searches introduced errors of omission and accidental inclusion of cases that searches by tested computer software would not be subject to. Data integrity was difficult to improve due to the large number of cases.

There was no "what-if" capability for scientific studies. Changing the sampling criteria amounted to beginning another study. This increased the costs of some studies substantially, while forcing a decision against changing these criteria in other studies for which extra funds were not available.

Soil researchers in Montana have had difficulty finding pedon data collected previously from their study areas or from similar sites. Their only recourse was to recompile the data. This may have meant resampling and analysis of pedons at considerable cost. A complete pedon characterization may cost $1500.00. Some of these data are used only once and then shelved. Soil Survey teams in Montana routinely discard pedon description data once the Survey of the area is published. These data should be used for research long after the original objectives for their collection were met. This requires a standardized system with as few steps as possible leading to the incorporation of these data into a permanent database.

The objectives of this thesis study can be divided into two major categories. System objectives covered the actual development of computer software and documentation. Applications objectives concerned the various uses that can be made of the database as a part of the Soilgrasp System.

System Objectives

1. Recode and rearrange the existing soils data sets such that all have identical formats and can thereby be joined to form a single database.
2. Standardize the variables and file formats of existing soils data in Montana to be upwardly compatible with the

current Soil Conservation Service national soils databases.

3. Implement a database management system (DBMS) package which would fulfill Montana's needs for soils information delivery.
4. Design the overall System in a way that minimizes user learning time.
5. Provide documentation for the System which describes its configuration and provides examples of its use, written in a style easily understood by university soils faculty and graduate students with a minimum of experience with computer systems.
6. Evaluate the System according to the set of criteria given in objective 3 above; with respect to its user friendliness; and according to its performance as a soil science research tool.

Application Objectives

1. Use the Soilgrasp System to access data for selected A-horizons in Montana, to clarify relationships of soil texture (particularly percent clay) and climatic factors with organic carbon content of soils in Montana.
2. Use the system to quantify the textural and some chemical characteristics of soils over several rock types in Montana, and to compare these properties for evidence of their inheritance from the underlying rock.

3. Use the System in a classroom situation, to enable graduate students to conduct original research with existing soils data for class projects.
4. Use the System to retrieve data for scientific research or extension agency use, according to specific sampling requests.

LITERATURE REVIEW

Database Management

The accumulation of environmental resource information has increased to such a level that manual data searches, though still performed, are impractical and expensive means of access. Heil (1982) stated the problem cogently:

> "There is universal agreement among natural resource planners that one of their major problems is finding and organizing data that is relevant, valid and reliable for use in decision-making.
> "One method for approaching this problem involves the application of database technology to the development of computerized information systems. Database management technology provides the mechanism through which data can be stored, integrated, and evaluated in many ways so that information can be easily and efficiently delivered to a diverse user community."

The large amounts of soils data already stored on computers are of little value unless they can be used to help us recognize problems and solve them. "The central problem of much data processing is extracting from a mountain of facts an essence of value to human users" (Martin, 1983).

Researchers and administrative decision-makers could gain valuable support from database management tools, but are often unaware of available technology (Priest, 1982). Databases and their management software need a higher profile in the workplace. The key to increased use of information delivery tools is a wider user base, including the end users of the data. The business field is far ahead of natural resource managers in this respect. Business database software has progressed to the point of using artificial intelligence techniques to make the computer appear to be an assistant. For the most part, resource managers still view the computer as something resembling a calculus quiz. Generalized (or subject) database management software can actually organize natural resource information as readily as that from any other discipline. There is seldom a need to wait for specialized applications software. Applications should be developed around the core database system.

Universities, state and federal agencies are actively exploring ways to organize extensive soils databases so that they can be integrated and shared. Most current efforts are in the experimental or prototype stages.

Montana State University has been involved in the development of soils information delivery systems since 1969, when work began to encode all available soil morphology and laboratory characterization data from the state (Decker, 1972). Decker's thesis work involved the testing of various data encoding techniques and retrieval methods, the actual entry of Montana's existing soil characterization data, and inventory and correlation of those data (Decker, 1972).

Decker envisioned and worked toward an Automated Data Processing (ADP) System for encoding and storing soils data. The objective for the system was to produce tables for land use interpretations based on soils information. Later additions to the system included intermediate data files and programs which could print out standard and popularized pedon descriptions, Soil Survey mapping unit information, and interpretation tables from the Soil Survey data (Decker, Nielsen and Rogers, 1975). They proposed extensions which would determine suitability and capability groupings for woodlands, range sites, windbreaks, and urban development.

The ADP System was difficult to use by anyone but those involved in its development. During the early 1970's, data were primarily entered using mark-sense forms and stored on keypunched cards and magnetic tape. All operations on the data were performed in batch mode on a Xerox Sigma-7 mainframe computer. Interactive (or on-line) access to the data was expensive and seldom possible. FORTRAN IV programs were written for each task, and were completely dependent upon the data file structures. This impaired their applicability and portability to other data sets. Data files were not stored on-line as hard disk files, but were to be loaded for each use from appropriate magnetic tapes. The file names were not recognizable to non-familiar users: program names were difficult to distinguish from raw data file names or those of interpretation table files.

The ADP programs that were developed assumed a data format compatible with a proposed NCSS coding system from 1973, whose format was later changed (Soil Survey Staff, 1979). Data sets encoded during Decker's thesis program were compatible with the 1973 codes. However, after that time, new soils data were not standardized to any specific format. Some data types, record lengths, variable codes and measurement methods were thus incompatible between data sets. An applications program written for one data set was useless for other sets. Expanding the sample size thus required rewriting parts of the applications programs. Few such attempts were made. It was clear that to increase the usage of Montana's soils information, several improvements were necessary. These included faster and easier access to the data, standardized file formats for all data files, the means to update the information, and enough documentation to allow new users to learn the ropes. Martin (1983) noted that if a terminal (or program) is confusing or difficult to use, it simply will not be used. Soils information system designers must make user friendliness a high priority, or even the most powerful systems will be shelved, with potential users opting for old, trusted methods.

Recent advances in database technology have been incorporated by the U. S. Army Corps of Engineers Construction Engineering Research Laboratory. This group has developed an extensive environmental data management network, one part of which is an interactive system called SIRS (Soils-5 Information Retrieval System) (Goran, 1983). The system accesses SCS-SOI-5 data, maintained by the Soil Conservation Service, and displays land use interpretation tables on the user's screen. The system is housed at the University of Illinois at Champaign, but can be reached through phone lines from any terminal. Sampling and search capabilities are available. The output is designed to present generalized information about soil series and land use recommendations rather than raw data from actual pedons.

Another soils data system is under development by the USDA-SCS Information Management Services (IMS) group in Colorado. This database consists of over 30 computerized Soil Surveys, with a geographical orientation. The database has been managed for many years by System 2000, a hierarchical Database Management System. Recently, the IMS group has replaced that software package with a more versatile relation-style system.

Soils and Parent Material

Extensive areas in southwestern and central Montana contain soils developed from limestone, intrusive igneous and extrusive igneous parent materials, or rock types. The degree to which the properties of these rock types are retained is greatest in young soils with cool and dry climates (Veseth and Montagne, 1980; Birkeland, 1984). Soil reaction and pH may reflect the composition of the soil's parent material (Ehrlich et al., 1955). Harradine and Jenny (1958) reported that under identical climatic conditions soils derived from basic (extrusive) igneous rock consistently had finer textures than soils derived from acid (intrusive) igneous parent material. Textural variability was higher and moisture equivalents were lower in the acid-igneous soils than in the basic-igneous soils. Soil texture in some Pennsylvania soils is inherited from parent material (Ciolkosz et al., 1979). Parent material texture affected weathering rates and subsequent soil properties in Scotland (Smith, 1962). Kadeba (1978) noted that parent material influences the relationship between percent clay and soil organic matter content in some regions.

Yaalon (1975) discussed the feasibility of separating parent material functions, or lithofunctions, from the other soil-forming factors considered by Jenny (1941). These factors include climate, topography (relief), biological activity, and time. Yaalon concluded that of the five factors, parent material is the most difficult to consider quantitatively as an independent factor because its effects are masked by those of the other, more dynamic pedogenic processes.

Organic Carbon in Montana Soils

Organic matter is an important soil constituent but factors that control the amount of organic matter in soils are not fully understood. Researchers have related accumulation of organic matter to factors influencing soil characteristics such as aspect (Losche, 1967), soil temperature (Anderson and Coleman, 1985; Jenny, 1941), and elevation. Anderson (1979) reported that precipitation and soil moisture regime also influence humus formation in Canadian grassland soils. Cool soil temperatures, anaerobic conditions and high water tables associated with depressions in the landscape can cause increased accumulation of organic matter. Similarly, organic matter in the Nigerian savanna, where precipitation controls plant growth, is closely correlated with the number of rain days per year. Latitude, parent material, soil clay content and human intervention, through cultivation and burning, also influenced the organic matter status in those soils (Kadeba, 1978).

Cannon and Nielsen (1984) reported positive correlations between grass production and both elevation ($r = 0.44$) and mean annual precipitation ($r = 0.53$) in the Northern Great Plains of North America. Organic matter percentage in the mollic epipedon showed even higher correlations with elevation and precipitation, with r-values of 0.62 and 0.71, repectively.

Since fine silt and clay particles combine with organic materials to form stable complexes (Kononova, 1966), soil texture is also expected to influence soil organic matter contents. In some regions, as much as 50 percent of the soil humus is held by clay particles in organo-mineral complexes and is thereby protected from rapid decomposition (Anderson et al., 1981). In the Southern Great Plains states, clay percentage was the major factor correlated with organic carbon content, and precipitation was also related (Nichols, 1984). McDaniel and Munn (1985) grouped 143 Mollisols and Aridisols in Wyoming and Montana into three temperature regimes (cryic, frigid, and mesic), and determined that texture is significantly correlated with organic carbon content only in the mesic soils. McDaniel and Munn used a weighted organic carbon content to a depth of 40 cm in their study, whereas Nichols (1984) used only Al-horizon data.

MATERIALS AND METHODS

Updating the Database

In September of 1983, the decision was made to recode and rearrange the existing soil pedon data files for Montana to conform to the latest national standards released by the Soil Conservation Service. A copy of the revised Pedon Coding System for the National Cooperative Soil Survey (NCSS) (Soil Survey Staff, 1979) was obtained. This publication clearly outlined the sequence and data types of all variables within the system. The exact location and column width for each variable were given. For any variable containing codes instead of real numeric or alphanumeric data, each code or combination of codes was defined.

The state's soil pedon data had been split alphabetically and by year into 31 files and stored on magnetic tape. Two sets of data existed: pedon description files which included site, horizon morphological, and laboratory characterization data for 270 sites; and Official Series Description files that lacked laboratory data but were otherwise similar, for 567 sites. Each case represented a single horizon.

Due to the absence of documentation for these data files, their complete formats were unknown. Four weeks of comparing these huge files to find repeating elements and patterns revealed that each case occupied not one but three records of a data file. A single column embedded among 3509 repeated the sequence from "1" to "3" down the file. Also, the record lengths were found to be variable, not fixed. Aside from these inconsistencies, the formats matched that of a proposed NCSS pedon coding system released in 1973. A revised version of this format (Soil Survey Staff, 1979), chosen for the updated database, contained a number of column and code changes. An exhaustive comparison of the two formats was made to find every change. FORTRAN 77 programs were written to read in the old data files 3 lines at a time and output new files in which each case occupied a single record. Further programs were written to read each record, change appropriate variable positions and codes, and output new, updated records. At the end of this process, which took approximately two months, the files were standardized and ready for a database management system (DBMS).

The new soils data files were merged to form a single master file. Extensive error-checking was performed on the first 10 cases (by hand, since their record length was too large for access by computer word processors). A single mistake in formatting one of the 1016 variables would have offset data for all variables beyond it. A few errors were indeed found and corrected.

Two smaller data sets at Montana State University contained pedon description information which could be used to expand the database.

The first data set was derived from a Bureau of Land Management study of Montana's native range sites (McDaniel et al., 1982). It contained data for 167 sites (763 horizons, or cases). This data set had been maintained by GRASP (Generalized Rapid Access Software Package), which is discussed in a later section. Many variables concerning ground cover percentage and production of native plant species were extraneous to the NCSS Pedon Coding System. These variables were included in Montana's soil pedon database, which thus became a logical superset of the NCSS Coding System while maintaining compatibility.

The BLM soil variables required recoding and rearrangement into the new database format. Variables without data were blank-filled so that records could be of fixed length. The column width of vegetation cover-presence variables was reduced from 7 to 3 columns to minimize database storage requirements. Approximately 4 Megabytes (million bytes, or characters) of storage space were conserved by this action, which simply eliminated unnecessary blank spaces. Likewise, shrinking vegetation production variables from 7 to 4 columns saved 5 Megabytes of storage space.

A second data set was compiled by Thomas Burke for his thesis study (Burke, 1984). This data set included new variables concerning soil moisture content to several depths, crop type, and crop yield. It contained data for 184 sites. All new variables were added to the main database and the soils information was recoded and rearranged as described for the BLM data set.

In anticipation of future database needs, new variables for fertilizer application levels of the plant nutrients N, P, K, Ca, B, Fe, Mn, S, and Zn were created within the database.

Three compatible data sets existed after the recoding process. These three sets were merged to form a new master file, "MT_MASTER".

GRASP: The Database Management System

The selection of an appropriate database management system (DBMS) for Montana's soil pedon database was based upon a number of criteria. The DBMS had to:

a. manipulate soils data of all necessary data types;
b. handle an unlimited number of cases;
c. accommodate a large number of variables, in the range from 1000 to 2000;
d. sort through the data rapidly and at minimum expense to users- certainly faster, cheaper, and more easily than performing the same tasks manually;
e. be used to update and maintain data, with simple commands for editing and deletion;
f. be accessible to soil scientists throughout Montana, and elsewhere in the United States if it became necessary; and

g. communicate with other hardware and software such as geographic information systems, microcomputers and state-of-the-art statistical software.

GRASP was selected as the DBMS because it fulfilled most of the criteria well. GRASP is a set of FORTRAN programs written by the staff of the Center for Data Systems and Analysis (CDSA), a unit of Montana State University in Bozeman. They are housed on campus in the Renne Library building. GRASP is still under development at this writing and is not available for purchase. GRASP runs on MSU's Honeywell Level 66 DPS/C3 mainframe computer and is dependent for certain procedures upon that computer's CP-6 operating system. Inferences in this document to CP-6 refer to the Honeywell computer: "CP-6" is the name used at MSU to mean both the operating system and the hardware.

A full analysis of GRASP's performance according to the above criteria is given in the Results section.

Database Inversion

Preparations began in April of 1984 to set up the database for GRASP use. GRASP required a list of variables for its inversion process, which it would use to flip the two-dimensional database file so that each record contained all the data for a single variable.

The variable list, termed an "invert file", would be used by GRASP as the DBMS built its own "directory file". Each entry in the list contained a unique 12 (or fewer) character variable name; a data type declaration; beginning and ending column numbers which represented the position of that variable in each record of the master (uninverted) data file; and a 40 character variable definition.

It was after completion of the invert file that the column widths of 415 vegetation variables were reduced as described above. A FORTRAN 77 program called "SHRINK_VARIABLES" was written to automatically update all of the invert file entries affected. A similar program, "COLUMNS", corrects column numbers of variable entries after the deletion of a variable, or when a mistake has been found. These programs are listed in the Soilgrasp System User Manual (Appendix) as models for future updates.

The invert file was scanned by GRASP for syntactical or typographical errors prior to data inversion by inserting a "NORUN" command near the end of the invert file. GRASP noted several typographical errors, which were corrected. The inversion was subsequently run with a subset of cases to check for formatting or data type errors. When this step revealed no errors, the database was ready for inversion.

GRASP can only invert 500 variables at a time. The invert file (MTINVERT) was separated into four smaller files, MTINVERT1-4. These files would create four data files when run in sequence. Site and horizon morphology variables were inverted first. During this run GRASP initiated the directory file, MTDIRECTORY, and created the first of its own data files, MT_SIHZ.

Three subsequent inversions created data files called MT_LAB, MT_VEG, and MT_YIELD. Each time, GRASP added to its directory file the information it needed to locate the data for all new variables.

Database Storage and Backup Procedures

All of the GRASP data files and the database directory were stored together in four different places. One copy was set up permanently on hard disk to allow interactive or batch access to the information at all times. Each backup copy of the complete database occupied two magnetic computer tapes. One of these backups is stored on rented labeled tape at Montana State University's Computing Services Center. A charge of $1.00 per month is assessed for this service. Another backup set is stored on user-owned tape in MSU's Plant and Soil Science Department. The third backup set is stored in MSU's Statistical Services office.

The file, "MT-DATABASE-BACKUP", was used to perform the backup procedure onto rented magnetic tapes 0819 and 0911. Its commands are listed below:

```
!JOB ISS005,NIELSEN D(00:05,1/3/86)
!RES MT=1,TIME=3:00:00
!DIRECTORY DP#SOILS.ISS005
```

```
!TAPE(0819/0819/WE)
!PCL
SEVERITY ABORT
MOUNT LT#0819 RING
C DP#SOILS/TAPE0819DIR.ISS005 TO LT#0819/TAPE0819DIR
C DP#SOILS/MTDIRECTORY.ISS005 TO LT#0819/MTDIRECTORY (XTE)
C DP#SOILS/MT_SIHZ.ISS005 TO LT#0819/MT_SIHZ (XTE)
C DP#SOILS/MT_LAB.ISS005 TO LT#0819/MT_LAB (XTE)
REW LT#0819
L(A) LT#0819 TO DP#SOILS/LT0819LIST_BAK.ISS005
REMOVE LT#0819
MOUNT LT#0911 RING
C DP#SOILS/TAPE0911DIR.ISS005 TO LT#0911/TAPE0911DIR
C DP#SOILS/MT_VEG.ISS005 TO LT#0911/MT_VEG (XTE)
C DP#SOILS/MT_YIELD.ISS005 TO LT#0911/MT_YIELD (XTE)
C DP#SOILS/SOILGRPROG.ISS005 TO LT#0911/SOILGRPROG (XTE)
C DP#SOILS/MT_NEWVARS.ISS005 TO LT#0911/MT_NEWVARS (XTE)
REW LT#0911
L(A) LT#0911 TO DP#SOILS/LT0911LIST_BAK.ISS005
REL LT#0911
C DP#SOILS/LT0911LIST_BAK.ISS005 TO LP
C DP#SOILS/LT0819LIST_BAK.ISS005 TO LP
END
```

To subsequently copy the database files onto the user-owned backup magnetic tapes described above, the terms LT#0819 and LT#0911 were changed to reflect the label numbers of each backup set of tapes. The system command,

```
!BATCH MT-DATABASE-BACKUP (SCAN)
```

submitted the file to be run at the deferred time and had the computer scan it for possible errors before execution.

A hard diskpack was rented for storing the database system files. This diskpack is stored permanently in the room that houses the CP-6 computer at Montana State University. It must be mounted each time a user needs to access files stored upon it.

Testing GRASP's Functions

At the time of implementation, GRASP was still under development. Each function or command was tested using the soil pedon database. If a command did not work as documented in the GRASP Users Manual (Allard, 1981) the CDSA was contacted. They quickly reported back with solutions. Notes were written in the GRASP Users Manuals for future documentation changes. We especially wanted to be certain that GRASP could manipulate character (alphanumeric) data such as soil series names. We had to be able to sample by such character strings, eliminating undesirable series or including selected ones in the sample. We also tested GRASP's ability to handle very long string variables (up to 100 or more characters), since site location descriptions were encoded as "notes" which occupied 120 columns of the database.

SOILGRASP: A User Friendly Interface

The CDSA was working on an applications database system for molecular compounds which had a menu-driven front end. A front end is a program that interacts with users, giving them several options rather than a prompt symbol. Work began in December, 1984 to convert a copy of this FORTRAN program to serve as a GRASP front end interface. The new version was given the name SOILGRASP.

SOILGRASP accesses menus stored in another file and shows appropriate menus to the user during a session. The menus were completely rewritten by the author to show choices based on GRASP options. SOILGRASP was designed to build a file of input commands which directs GRASP to perform the series of operations requested by the soil scientist through these menus. The choices offered include most of the common commands used in soils studies.

Improving Data Integrity

Data for the elevations of 65 sites within the database (none of which had elevations recorded) were obtained from topographic maps. With a repeated series of sampling and editing commands these data were entered into the database.

After the Soilgrasp System was built, all cases from the BLM data set were found to be encoded using decimal points where needed for the variable "PH_VALUE". All other cases in the database had been encoded with a single implied decimal place, with no decimal point actually entered. The variable was inverted with the declared data type "REAL(1)". This type indicates an implied decimal point with one decimal place. When GRASP inverted the data for pH, it automatically divided by 10. BLM pH values were thus one tenth of their original value. These cases should have had a "REAL" data type. Since the error was global for BLM cases it was possible to edit them all with one set of commands:

```
?SELECT: BLM_CASES
?EDIT PASS=GRRR: PH_VALUE = PH_VALUE*10
```

A special bit variable, "BLM_CASES", had been created which could be used to select for these cases alone. When the editing was finished, a listing of all affected cases was checked for accidental inclusions.

Township and Range location information was not recorded for cases from the original state pedon data set. This information could be valuable if the Soilgrasp System was ever used to provide data to a geographic information system. A file containing locations for these cases was created by work study students from the "location notes" of the original data files. The locations would all be inverted into the database as new variables. At this writing, the latter had not yet been performed.

Soilgrasp System Documentation

The name Soilgrasp was given to the completed System, a name derived from that of the interface program. Common operations performed with the System are described in the Soilgrasp System User Manual (Appendix). This Manual is written in language that users with minimal computer experience can understand. A glossary is included, as well as an annotated list of references which may be read for a greater understanding of these systems in general.

Using the Soilgrasp System

With the basic components of the database system linked- the full database, the management system, and the user interface- its variety of functions could be addressed.

The first applications were simple attempts to extract data summaries from the database, both to check for data integrity and to answer basic queries about Montana soils. A demonstration was given for representatives of the state Soil Conservation Service office. Sites with crop yield data were selected and sorted by ascending yield. A listing of the site keys, soil series and yields was then printed out as a ranking of series by crop yield. In other example studies, lists of site characteristics, soil chemical properties, and classifications were produced. Soil horizons were chosen using increasingly complex sampling criteria. The sampling procedures used the logical operators AND, OR, EOR, and NOT to combine expressions that defined variable ranges. Relational operators (such as greater-than and less-than) were included, as well as mathematical formulae which transformed the data before case selection.

A listing of soil texture data for A-horizons on cultivated lands in Montana, and basic statistics for each