THE CANADA SOIL INFORMATION SYSTEM (CanSIS)

GENERAL USER'S MANUAL

K. B. MacDonald and B. Kloosterman

Land Resource Research Institute
Ottawa, Ontario
LRRI Contribution No. 83-59-E

Research Branch
Agriculture Canada
1984

Copies of this manual are available from:

Land Resource Research Institute
Research Branch, Agriculture Canada
Ottawa, Ontario, Canada
K1A 0C6

Produced by Research Program Service

TABLE OF CONTENTS

APPENDICES

List of Tables and Examples

List of Figures

ACKNOWLEDGEMENTS

In this manual, the authors have attempted to incorporate the many valuable comments and suggestions from land resource scientists and technicians across Canada. While the authors accept full responsibility for the material included and its interpretation, they would like to thank the members of the land resource research community for their continuing support, interest and patience.

The specific contributions noted for the following individuals are gratefully acknowledged. The sections on RAPID data dictionaries and on Automated Coding Procedures - The Computerized Legend were developed and written in their original form by Miss C. Atkins. The Procedures and Reports using Standard Software Packages were developed and written by Mr. Z. Strezelczyk. Miss V. Kirkwood and Mrs. P. Benson (Parks, Environment Canada) wrote and rewrote the Data Retrieval Section on DREAM. The typists, S. Parent, K. Burns and M. J. Berriault transcribed, corrected and reorganized the fragmented sections. Finally, Mrs. B. Lacelle wrote the section Soil Names File Data Organization and competently carried out the thankless tasks of synthesizing and incorporating the reviewers' comments and guiding the manuscript through final typing, translation and publication.

To the people who reviewed the manual, P. Benson and W. Fraser (officially) and the correlation and cartogrphic staff of LRRI (unofficially), your suggestions have improved the content and readability of this manual. Lastly, to the systems analyst staff of CanSIS - P. Brimacombe, B. Gronas, and L. Boardman - a special "thank-you" for your unceasing efforts to keep us honest and accurate.

General Users Manual

1. INTRODUCTION

CanSIS, the Canada Soil Information System, is a computerized information system. It was developed and is maintained by the Land Resource Research Institute, Research Branch, Agriculture Canada, in cooperation with provincial research agencies and universities with additional support and input from Parks Canada, Department of the Environment. The techniques, procedures and computer programs which make up CanSIS are well suited to the storage, manipulation, and display of data related to soil science and natural resource studies. CanSIS is comprehensive in design and concept, and operates at a national level to provide a unified framework of data organization under which soils, agronomic and other resource data can be stored and manipulated.

2. OVERVIEW OF CANSIS GENERAL USERS MANUAL

2.1 Users of CanSIS

In broad terms, the users of CanSIS fall into two categories "information" users (readers) and "system" users (authors).

The information users represent people or agencies who require access to basic soils, agronomic or natural resource data. In this context, they are primarily users of data; they may or may not be collectors of data.

In the past, these data were published in research reports and scientific publications. The data are still available in these "hardcopy" forms but now they are also stored in a standardized fashion in CanSIS. The implications of having the data in CanSIS are that the "information" user can now search in one single computerized archive to determine whether the required data are available. If they are, he can use the computer to manipulate these data and prepare tables, charts, and carry out statistical analyses specifically tailored to his application.

The "system" users represent persons or agencies who are involved in collecting, organizing and managing data related to the soils, agronomy or resources. For this group of users, CanSIS represents a system which provides most of the capabilities required to manage and store their data. By using CanSIS, they do not have to spend time and money redeveloping existing capabilities. CanSIS represents a system with a high degree of standardization which means that by using the capabilities already in place a user has facilitated comparisons between his data and similar data collected by other user groups.

2.2 Purpose

The purpose of this manual is to provide a general definition and description of CanSIS and, in addition, to document or give reference to the more detailed technical procedures required to use CanSIS. The specific objectives of this manual are:

(i) to provide a general overview and description of CanSIS
(ii) to specify the procedures to be followed to input data to CanSIS
(iii) to document the techniques required to access data stored in CanSIS

For these objectives the primary focus will be towards data files of national interest which already exist in CanSIS, especially those designed for easy access by regional users. Reference will be made to the CanSIS capabilities for specific project files and for those files which are more complex in organization and as a consequence require more specialized procedures to access data.

The manual is also designed to provide a general description of the types of data files stored in CanSIS which are of national interest. For these national files the final responsibility for file maintenance and data quality resides with the CanSIS staff at LRRI.

2.3 Scope

This CanSIS User's Manual includes a general description of procedures, techniques and capabilities of CanSIS (as it exists in 1983). Detailed descriptions and specifications of procedures which have already been documented are referred to but documentation is not duplicated. This statement applies particularly to the larger more complex procedures for which separate manuals already exist. These procedures are listed below and the complete reference for the associated documentation is given in Appendix 1:

Capabilities Documented by CanSIS Staff

1) Input Manual for Soil Data File (Detail)
2) Output Manual for Soil Data File (Detail)
3) CanSIS File Management System
4) Input Manual for the Soil Performance and Management File
5) Output Manual for the Soil Performance and Management File
6) User's Manual for the Land Potential Data Base
7) CanSIS Cartographic File
8) Spatial Display Point Plot Package
9) Computerized legends: Development and Use
10) Wetlands Registry Manual
11) Accessing CanSIS from an external computer account
12) CanSIS File Retrievals

Capabilities Documented Commercially

1) IBM Datacrown Service Guide
2) Easytrieve
3) SAS (Statistical Analysis System)
4) Rapid

For each data file of CanSIS specific procedures have been developed to input data, manipulate and retrieve them.

Specifically each data file of CanSIS requires;

1) a data collection form and instructions for its use
2) a description of the steps and procedures required to input data from a completed form into the computer for Retrievals:
3) a detailed description of the fields of data and their organization within the computer to describe the format in which the data are available for access
4) a set of procedures allowing the user to select the data which are pertinent to his requirements, and to display the information[1] on a cathode ray tube or print it at a "hard-copy" terminal for a quick scan of the data to determine if the file contains a sufficient quantity of data of interest to warrant further output.

[1] In addition to the documentation contained in this report, the user must obtain logon procedures specific for the installation he is using. These can be obtained by contacting the CanSIS project in the Land Resource Research Institute. A typical example of logon instruction is illustrated in Appendix 2.

5) more complex output reports and data manipulation achieved by using commercial software packages such as EASYTRIEVE (primarily for report writing) and SAS (statistical analysis system) for data manipulation and statistical analysis. Examples of some standard types of reports developed using these packages and annotated listings of the programs are referenced in Appendix 6-3. A detailed discussion of the use of these packages is beyond the scope of this manual and the user is referred to the adequate commercial documentation supplied with them.

In general terms, the procedures and techniques used in the steps outlined above are common to most CanSIS files (except cartography). In this manual they are illustrated for the Soil Names File.

2.4 Organization

The main body of the text of the General CanSIS User's Manual contains information which is relatively descriptive in nature and of general interest. So far as possible this information should remain constant with time. Extensive use is made of appendices throughout the manual. These appendices contain the more detailed and technical information relevant to specific topics and also the details which are most susceptible to change. The start of each appendix includes the date of its preparation. Users concerned with the content of specific appendices should verify that they have the latest version. Where appendices describe properties of individual data files, the Soil Names File has been used as an example. Comparable appendices are available on request for other files, in some cases e.g. P/M and Detail they are included in manuals specifically related to those files.

3. DATA FILES OF CANSIS

Two levels of data files are stored in CanSIS - national files and project files. The national data files have been established to organize and store reference data on basic soil properties, agronomy and resource information. Within these files the data conform to nationally accepted definitions which are applied in a consistent fashion in all regions. The files are as comprehensive as possible to allow some data specific to particular regions to be included within a national framework. All data in the national files have been verified by the regional staff and also by CanSIS staff to the fullest extent possible; occasionally errors are detected by users of the data and are corrected as they are brought to the attention of CanSIS staff. The basic data in the national files are available to all qualified agronomists and soil scientists either in "hard-copy" print-out form or in machine readable form on magnetic tape.

For most national files, data are submitted on an ongoing basis by federal, provincial, university and industry agencies and researchers. The final acceptance of data for input to a national file is controlled by CanSIS staff in Ottawa. Data are input as they are received and stored in tape files but are not generally accessible in this format. Periodically, normally one to two times per year, they are reformated and stored on disk for on-line access and manipulation.

Data in project or experimental files are of interest to a limited group of users (generally the ones who collected the data). In many cases the data have not been thoroughly verified and analyzed. For these reasons, access to the data is limited to the owners of the data or persons designated by them. The data are managed within CanSIS to the extent that they are accommodated within the national file structure; the responsibility for additional manipulation and analysis is left to the user.

3.1 Established National Files

3.1.1 Soil Data File: This file contains field description and field and laboratory analysis of modal soil profiles. New descriptions and data are added on an ongoing basis as they are collected by federal, provincial and university personnel and subject to national correlation. The input of data is coordinated by the CanSIS project leader. The data are archived in a single file on tape, from which standard profile descriptions can be prepared. The data have also been reorganized for storage on disk to allow for on-line access. The data are represented in the following three RAPID relations and an associated RAPID monitoring relation.

(i) Site - describing the geographical location, landform, soil taxonomy, other site characteristics.
(ii) Morphology, Physical and Chemical Analysis - giving details on horizon depths, morphology and other field characteristics as well as the data from laboratory physical and chemical analysis.
(iii) Notes - containing descriptive and explanatory information.
(iv) Monitor - a relation identical in structure with the morphology, Physical and Chemical analysis relation but used for repeated measurements over time at a site. This represents a project type file within the existing file structure.

3.1.2 Soil Names File: This file contains a listing of all soil names and their acceptable abbreviations in use in Canada, their status (proposed, current, archived) along with characteristics and principal attributes of the soil. Data in this file are submitted by regional survey units for acceptance at the national level. The status is assigned by the national correlator. The data are archived on a single file and have also been stored on disk as a single RAPID relation.

(i) Soil Names-containing the soil name, province, abbreviation, status and attributes.

3.1.3 Performance/Management File: This file contains data on the yield, quality, growth and development of crops at recorded locations and under specific conditions of management, climate and soils. Data are submitted on an ongoing basis by federal, provincial, university and industry agencies and researchers. The final acceptance of data for input to the national file is controlled by CanSIS staff in Ottawa. In some cases performance/ management data are submitted for specific projects and managed in user specific or experimental files until they are released by the originators of the data. At this time they are screened for quality, aggregated from the replicate level to the treatment level as required and input to the national files. Data are stored in seven separate tape files but are not accessible in this format. They are reformatted and stored on disk for on-line access and manipulation. The data are represented in a series of 11 RAPID relations.

(i) Site* - describing the geographic location, landform, soil taxonomy, other site characteristics.
(ii) Horizon* - giving details of the horizon depths, morphology and field characteristics.
(iii) Soil Analysis* - the results of laboratory physical and chemical analysis.
(iv) Global Management - general land use, past years' management, management applied over the entire site.
(v) Factor - experimental factors or parameters under study.
(vi) Factor-Level - levels as a range of values within each experimental factor.
(vii) Treatment - definition of treatments as combinations of specific levels of factors.
(viii) Weather - records of above ground weather observations.
(ix) Crop Development - cardinal dates of crop development, damage and causes.
(x) Crop Yield - observations of crop yield and quality.
(xi) Notes - any descriptive, explanatory or interpretive information.

*These relations are similar in form and information content to the corresponding ones in the soil data file, but in many cases the information is not as extensive.

3.1.4 Wetlands Registry: This registry consists of five relations defined on the RAPID relational data base linked together by the common key elements province, year and area identification number. The distinguishing feature between the relations is the degree of specificity of the key. Data are submitted on an ongoing basis and

input is coordinated through the peatlands project in LRRI. The following relations are used for storing and manipulating wetland data;

(i) Area - a relation which describes all properties common to all sites within a specific identifiable wetland. The key consists of a designation of province, year, and area identification number.

(ii) Site - a relation containing all data specific to a site and having a singular value for each site i.e. horizons or layers are not described because they constitute multiple values. The key to this relation consists of all the fields from the area relation plus a designation of x and y coordinates about the geographic location of the area as specified in the area location.

(iii) Hydrology, Morphology Physical and Chemical analysis - this relation describes attributes recorded and measured in the field and in the lab dealing with specific depths and/or sample types. The key is specific to site, sample number (which is layer in most cases or type of water sampled) property being measured or described and date of sampling. The inclusion of date allows for repeat sampling of the same property with time. In addition, the relation has a secondary key of attributes to facilitate grouping and analysis. Finally the most important attributes are listed in specific fields to facilitate their use for comparison and analysis. This redundancy is planned for convenience.

(iv) Vegetation - This relation records details about the vegetation present at the site. The key consists of the key fields from the site relation plus the designation of class and number.

(v) Notes - A relation containing all free format observations recorded. The key consists of the key fields from the site relation plus date (day, month) and note type and a line number.

3.1.5 Cartographic File: Contains line and symbol information from digitized (computerized) maps. Maps in manuscript form are submitted to the Cartographic Section of LRRI after the maps and legends have been finalized and edited by the Agriculture Canada correlators. Data are input by the cartographic staff and stored on tape. Each map is stored as a separate file on tape.

The hardware and software associated with this subsystem are relatively complex and have not been made accessible to users. Users specify their requirements for derived or interpretive maps via computerized legends or coding documents and the maps are prepared by CanSIS staff in Ottawa.

3.1.6 Computerized Legends - The computerized legend contains the information in the map symbol, the number and extent of map areas represented by each map symbol and any additional interpretative data which can be associated with the map symbol.

Data in computerized legends are organized into one or a series of RAPID relations. They facilitate manipulation and analysis of information associated with the map and preparation of derived maps.

Computerized legends can be prepared for any digitized map or series for which there is sufficient interest. Appendix 13.1 lists the computerized legends currently available.

3.1.7 Land Potential File: This file represents an extended computerized legend for the Soils of Canada map (1:5,000,000). In addition to the usual information associated with a map symbol (soil taxonomy, region, parent material, etc.), this file contains data on climate normals, agricultural capability and estimates of yield and potential yields for a variety of crops. This file was prepared during the course of a project at LRRI. Further details on the methods by which the climate data were prepared or how the crop yields were estimated are given in Modelling Methodology for Assessing Crop Production Potentials in Canada (1981) by R.B. Stewart, Technical Bulletin 96, LRRI Agriculture Canada and Crop Production Potentials for Land Evaluation in Canada (1981) by J. Dumanski and R.B. Stewart, LRRI Agriculture Canada.

The data in the land potential file are stored on disk in a single RAPID relation.

3.2 Project Files

Project files are experimental or user specific data and contain information which is collected by a specific group of users. So far as possible these users attempt to format their data into forms compatible with similar data in national level files. In some cases the data are stored at a greater level of detail (e.g. crop yield and quality data collected at the individual replicate level); at other times they represent special types of data which have no national counterpart (e.g. the files developed for Parks Canada to store wildlife observations made by park wardens); in many cases the data are collected in conjunction with a specific project or soil survey and are used to refine and confirm the map delineations (e.g. the soil dailies). Generally these records consist of a restricted portion of the information normally collected for a record in the soil data file.

For all these cases the data are of interest to a limited group of users (generally the ones who collected the data). Access to the data is limited to the owners of the data or persons designated by them.

3.3 Establishing New Files

Many of the standard CanSIS capabilities and techniques will apply in cases where new files are to be established but some additional development is required.

The steps associated with the establishment of a new file are summarized in Table 1 (and described in detail in section 10). For users who are operating within the definitions of existing files these steps are not required. Certain sections such as design of the data collection forms, and keypunching instructions may be relevant if the standard data collection form is to be modified. Users are strongly advised that adapting existing files to new applications is generally easier and quicker than developing new files.

Table 1. Procedures for Establishing New Files (other than Cartographic).

1) Purpose of file.

2) Type of output desired.

3) Design of the Data Collection Form.

4) Data Coding Instructions.

5) Keypunching Instructions.

6) Definition of the data type and format to the computer by means of standard (CanSIS) procedures.

   a) for input, edit and update
   - to a tape file
   b) for manipulation and output
   - from a tape file to a disk file.

NOTE: These steps must be completed prior to the input and manipulation of data. Users are advised that these procedures can be long and time-consuming and should adapt existing files or portions thereof to their application where possible.

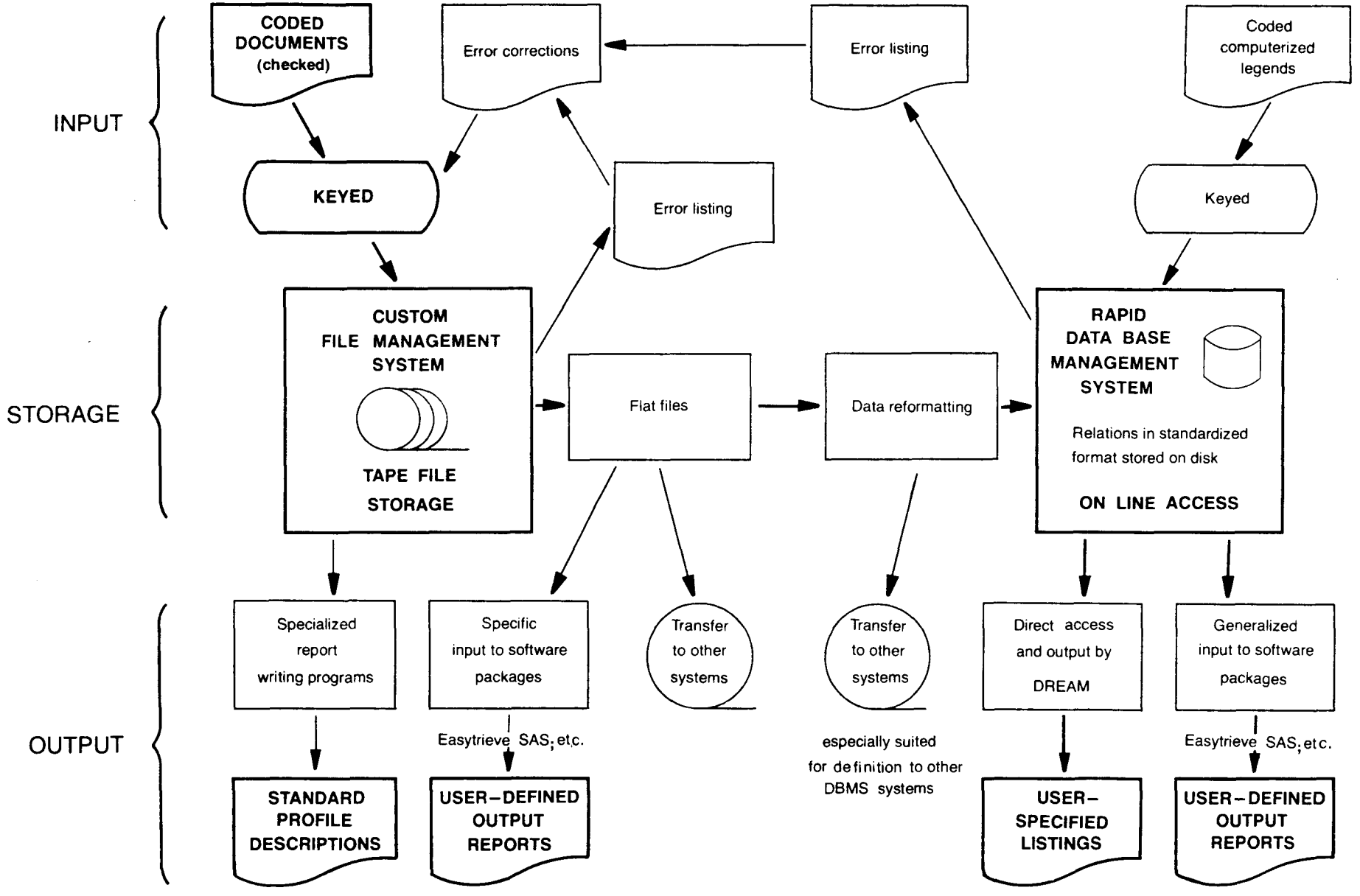


Figure 1. Generalized organization of CanSIS (for all data except cartographic)

Table 2. Flow of Data From Collection to Computer Output (data other than Cartographic).

1) Data Collection

2) Verification.

3) Keying of data
   - transfer to machine readable form.

4) Data Input and computer check for errors.

5) Error Correction.

6) User specified consistency checks.

7) Correction of inconsistency errors.

8) Reformatting the data for output file.

   a) for specific report e.g. profile description
   b) for definition into a disk file as part of a RAPID relation or into other software such as SAS.

4. OVERVIEW OF CANSIS DATA MANAGEMENT

4.1 Data Storage Format Considerations

In any computer system, information can be input and stored on magnetic tape or on disk. Tapes are a cheaper storage medium but must be mounted (with a cost and time involved) each time any data on the tape are to be accessed. Each time data on a tape are to be used the tape must be spooled from the start to the point at which the data are stored and then rewound. (i.e. data can only be accessed in sequential fashion). Tapes provide an efficient storage medium for large volumes of data which are accessed infrequently. Magnetic disk provides a storage medium which is significantly more costly than tape. It is constantly mounted and accessible on the system. Data stored on disk can be accessed in a random fashion. This facility is a definite advantage, especially when accessing a variety of data sets which are not in the order required. Disk storage is efficient for relatively small quantities of data which are accessed frequently. For greatest efficiency, data are shifted from tape to disk for periods of active use and archived on tape during periods of intermittent use.

All data in CanSIS are stored on magnetic tape for secure archival purposes. The data stored on tape exist in a variety of formats, many of them developed specifically for CanSIS. In general these data are accessed only with the assistance of a programmer analyst. This is generally done by the CanSIS staff in Ottawa. Some data which are accessed frequently are also stored on disk to allow immediate, random access. All data maintained on disk have been organized into a single format type. This has been done to facilitate access by non-system users to all the various types of data available through the use of a single set of procedures. At this stage where data have been reformatted, they can be easily transferred to other computer systems.

The transfer to other systems can be done most efficiently by preparing tapes containing the data in the disk storage format.

4.2 Flow of Data from Collection to Computer Output-The Data Subsystem

One of the major differences between the "system" user who inputs information and the "information" user who deals primarily with retrieving data is that the data are input into one or a series of files on tape. Any errors detected are corrected and the records on the tape are updated. The "information" user deals exclusively with files which can only be read but not altered or updated and for most files which are accessible on-line via a terminal, these data are stored on disk.

The "system" user must then deal with procedures for inputting data, editing and updating it, working primarily with files on tape. He must then deal with the procedures for reformatting the data from each of the various tape storage formats into the single disk storage format and then transfer the data into a series of standard disk files for manipulation and analysis.

In CanSIS most of the data input, update and error correction procedures have been established to operate with tapes as the storage medium. This

is done within the framework of a series of software programs which make up a file management system (FMS). The file management system was custom designed and developed for CanSIS. After the data are input and the errors have been corrected the tape files constitute a documented storage format which is defined and standardized for each file.

For some applications, "system" users of CanSIS may want to take over some or all of the data input procedures. Some of the reasons for this are:

(i) The CanSIS staff in Ottawa cannot provide the rapid turnaround required for some projects.
(ii) The user group collecting the data can edit it and correct errors much more easily and efficiently than the CanSIS staff.
(iii) In several cases the data are being used in CanSIS as it exists in Ottawa and also in portions of CanSIS or similar systems in place in the regions. In these circumstances the regional user must take responsibility for the interface between systems and can generally organize specific data transfer procedures.
(iv) For specific projects, it may be necessary to use the techniques and procedures of CanSIS to develop new information files and/or capabilities.

The CanSIS staff ini Ottawa is too small to support all these needs. Consequently, it falls to the specific project or region to develop the capabilities required.

The general organization of CanSIs and the flow of data through the system is summarized in Figure 1.

The steps involved in transferrring data from the original coding form into a computerized form suitable for output and analysisd are common to all types of data. These steps are summarized in Table 2. The data collection step and the author's verification are obviously carried out by the originator of the data. The remaining steps or some combination of them may be carried out by the user if circumstances warrant and if support is available or the user may simply be a passive "system" user and leave all subsequent steps to the CanSIS staff in Ottawa. With this latter approach the user's satisfaction with the system is limited by the size of the CanSIS staff whereas if the user is in control he will achieve the capabilities he wants within his own limitations of time and funds.

Once the data are clean, they are organized into a series of relations in a generalized data base management system called RAPID. RAPID was developed within Statistics Canada as a common solution to many statistical data processing applications. It allows for on-line iterative manipulation and selection of the data by the general user and easy interface of the data into commercial statistical software packages. (For more information on RAPID see sections 7 and 8.)

4.3 Access to CanSIS Data Records

Data stored in CanSIS are made accessible in the form of a number of relations in a data base management system called RAPID (a relational data base management system developed by Statistics Canada) described in section 7.

Because the same structure is used for all files or relations, the same means of access and reporting can be used regardless of the data. This means that the techniques to access and manipulate data described in this manual apply to a broad range of data stored in CanSIS, including map legends, soil descriptions etc.

In their simplest form the access procedures allow any user (a user with no previous computer experience could be trained to this level in approximately half a day) to determine what quantity of data are available and whether or not they are likely to be of use for a particular application. Generally, if the data are adequate and of interest, further manipulation will be required. This additional manipulation can also be carried out by the user through the use of generalized programs for preparing reports and/or statistical summaries. For most of these the amount of training required is fairly small (for most users with little or no computer experience, training in the use of the first generalized package will probably require 3-5 days, subsequent software packages would likely make use of the same basic capabilities; consequently, training times would be shorter).

Users in any region of Canada can access data in CanSIS[1] via the DATAPAC[2] communications network. This allows anyone with a computer terminal and coupler[3] to access DATACROWN by dialing a phone number in the closest major city and specifying the DATACROWN computer address. In order to access data in CanSIS the user must either have permission and the appropriate accounting and password details from Agriculture Canada or the user must establish an account with DATACROWN and obtain the necessary cross-authorization to read data stored in CanSIS.

---

1 CanSIS is resident on an IBM 370 & 3033 computer currently located with DATACROWN Inc. 650 McNicoll Ave., Willowdale, Ontario M2H 2E1. DATACROWN is the computer service bureau which is providing computer facilities and services to Agriculture Canada under contract. Some CanSIS procedures have been duplicated on the VAX 11/780 computer to facilitate most cost effective use of computer resources by employees of Agr. Can.

2 DATAPAC is a Canada-wide universal shared intelligent network for data communications. It provides an access point in every major Canadian city through which users can access computer facilities, data bases, other users, etc.

3 A coupler is a device used to link a computer terminal to a telephone line. Commonly an acoustic coupler or modem is used to provide access over any public phone line. This set-up is limited to a relatively slow speed operation of 300 baud. Where faster speeds or more sophisticated printing facilities are required it is necessary to establish direct wire links between the terminal and the DATAPAC network via dedicated phone lines.

The procedure required to access CanSIS data and software at DATACROWN from an account outside of Agriculture Canada are described in an internal LRRI Report[4].

All data in the national files of CanSIS are accessible to individuals and agencies who require such resource data and who are qualified to use it in a valid fashion. In many cases the data are stored in relations on the RAPID relational data base management system and can be accessed directly and manipulated as desired by the user. Alternatively, if appropriate facilities or skills are not available the user may request the data in "hardcopy" form. These requests are priorized and filled by the CanSIS staff in Ottawa.

While the data are generally accessible, the ability of CanSIS staff to respond to requests for specific kinds of output is limited by the availability of manpower and computer budgets. In all cases standard types of output can be provided in sufficient quantity to allow the user to determine whether or not the data are adequate for his needs. If large quantities of data are required or extensive data manipulation is needed it may be necessary for the user to provide the resources (personnel and/or computer budget to carry out the task).

---

4 Systems operation at DATACROWN using an Account Independent of Agriculture Canada. Some considerations for clients of CanSIS to use procedures developed and maintained by Agriculture Canada and for using data compiled and stored by Agriculture Canada. Internal LRRI Report by K.B. MacDonald and A. Sunnak, 1982.

5. INPUT OF DATA TO THE CANSIS DATA MANAGEMENT SUBSYSTEM

To achieve a data base with the lowest possible level of errors, the number of data transfer steps from collection to computer input should be kept to a minimum. Ideally, anyone collecting information for input to CanSIS should record the data in the field directly in machine readable form. In most cases, the techniques for direct machine readable data recording have not been adapted for data input to CanSIS. Standard data formats have been defined so that the data originator can record information in a format from which it can be transferred (keyed or digitized) directly onto a computer input medium (e.g. cards, magnetic disk, tape etc.).

5.1 Data Input Formats for Tabular Data

For all national files and most project files standard data forms have been prepared to facilitate data collection in a format suitable for keying. Generally these forms are comprehensive and attempt to encompass the range of information which would logically be stored in the file to which they pertain. The complete computer input form represents the range of information types which have been defined to the computer. Consequently, if a type of data or observation is not specifically identified on the standard data form it cannot be input and stored in the CanSIS file associated with that form.

The collection and preparation of data for computer input (in particular input to CanSIS) requires strict adherence to coding rules and conventions. These are determined by the software used to manage the data. In CanSIS this software consists of a custom developed file management system (FMS) and the data base management system RAPID. The software determines both the general coding conventions and also the styles of data input.

5.2 General Coding Conventions for CanSIS

1. Alphabetical entries are left-justified and numerical entries are right-justified.

MIDAS
PHØSPHØRUS
NITRØGEN

4
2500
3000

2. Decimal places are indicated by a shaded area. If no value exists for a decimal place, a zero must be inserted.

600
300
1200

3. All header key fields must be filled in up to the card type.

5.2.1 Styles of Data Entry for CanSIS

For data which are input and archived in the FMS (these include the soil names file, soil data (detail) file, performance management file, and some project files) there are four possible ways of recording data. For data which are input directly to RAPID only the "fixed entry" style of data input is used. In the following description of possible data entry styles the specific examples refer to the forms used to record data for the soil data (detail) file.

5.2.1.1 Fixed Entry: Fixed entry fields are data entries of a fixed length at a fixed location on the coding form. Location, vegetation, special notes, and all laboratory data, are examples. These data are entered either by the use of codes or in free format. For laboratory data, trace quantities are entered as "-1".

5.2.1.2 Self-defining Entry: Data are entered in one of two ways:

(1) by circling an entry name (a letter followed by a series of numbers) that specifies not only the field but also its value.

For example, D817** describes the field "Present land use" and specifies the value "Abandoned farmland". Within any one field an attempt is made to circle the best choice; intergrade situations can be handled by circling two entries.

(2) by circling an entry name and indicating the horizon number in which a variable is found. This is accomplished by inserting horizontal bars (-) for one or more of the appropriate top six horizons (horizons 1 to 6), and the number 1 for one or more of the appropriate lower six horizons (horizons 7 to 12) in the appropriate box provided. If you want to indicate, for example, that horizons 1 and 7 are true for that entry, place a plus sign in the first box.

This style of data entry is used exclusively for profile descriptions.

5.2.1.3 Value-coded Entry: The data slot (consisting of 6, 12, or 18 columns or spaces) consists of a unique entry name (two-digit numeral) followed immediately by a fixed field entry. For example, 02 describes the field "Percent slope". Behind this, and in the spaces provided, the measured slope is recorded as a percentage.

5.2.1.4 Composite Entry: Composite field entries consist of combinations of the second type of self-defining entry with fixed field entries. It is used where both value and position of a variable are recorded. For example, L001 defines the field "Material composition" and the value "Leaves". An estimate of the percentage of leaves in the material, and the horizon or horizons in which the stated percentage of leaves can be found, are then recorded in the space provided. This style of data entry is used sparingly in the forms.

5.2.1.5 Free Format Entry: Data recorded in special notes and interpretative comments segments are in the free format occupying one space per character; the data can be numerical or alphabetical, and either with or without punctuation. The purpose of these segments is to provide space for recording data that cannot be accommodated elsewhere on the form.

In that the data are recorded in free format, these segments have minimal capability for computer search in the near future.

5.3 The Soil Names File - An Example of Specific Coding Instructions for a CanSIS File.

In addition to the preceeding rules and coding conventions, a set of specific definitions and instructions is required for each CanSIS data input form. These instructions ensure that the data are recorded in a consistent fashion by all users so that general interpretations can be made. The following specific instructions refer to the coding of data for the Soil Names File.

The input document for the soil names file consists of a one page pre-printed form (brown lettering printed on both sides of a 21.5 x 35.5 cm page). A copy of the form is included in Appendix 3.

5.3.1 Contents of the Soil Names File

5.3.1.1 Header Key: The first two items of data (fields) to be recorded on the soil names file are the province and the soil code. Together these fields constitute a unique 5 character "header key" which distinguishes this soil name record from all others in the file.

The header key information is completed as follows:

PROV - The information recorded in this field consist of a 2 character code representing the province in which the soil was described. The valid codes for provinces and territories are listed in Table 3.

SOIL CODE - The soil code consists of a three letter code which must be unique within the province and which represents the soil name. For example soil code MEL represents the soil name MCLELLAN in province 01 (British Columbia) and soil name MELFORT in province 03 (Saskatchewan). It should be noted that this same three letter code also forms part of the header key of the Soil data (DETAIL) file and facilitates cross- referencing of these two files.

5.3.1.2 Data: The remainder of the soil names file input data form is used to record the characteristics associated with the soil which is uniquely defined by the header key. Up to thirty-one attributes are recorded as follows:

SOIL NAME - Soil names of up to 20 characters are allowed to identify a recognizable soil. Each soil name must be unique for all of Canada. Consequently, all soil names should be submitted to the national soil correlator shortly after establishment for registration and clearance. The CanSIS staff, and the national correlator will assist the provincial units in defining an acceptable, provincially unique, three character soil code. It should be noted that for retrieval purposes a soil name of up to 20 characters can be retrieved but names longer than 16 characters cannot be selected as the basis of retrievals using the software package DREAM. For this reason it may be advantageous to restrict the length of all new soil names to a maximum of 16 characters. The minimum acceptable number of characters is six.

REPORT NO. - This field provides space to record a reference number to be associated with the soil name. Normally this will be the number of the soil report in which the soil is first described and/or described in greatest detail. The information in this field must be numeric with a maximum length of three digits and must be entered right-justified.

PROXIMITY TO U.S.A. - is a self-defining field as are all subsequent fields except for subclimate. This field is used to indicate soils which occur within one soil report width from the U.S. border i.e. within specified limit by encircling C801** or soils which occur further from the border by circling C802**.

For all self-defining fields except PHASE only one value may be circled (for the soil names file).

STATUS - indicates the level of acceptance of a soil name. "C811** reserved" indicates proposed soil names which have been approved by the national correlator but for which only limited characterization has been completed in the file. "C812** open for mapping" indicates accepted soil names which are currently in use and which have been completely characterized. "C813** closed for mapping" indicates accepted and published soil names which are not appropriate for use in current or future projects. Frequently, they represent soils which on new maps would be divided into two or more soils.

LEVEL - indicates the level of detail or concept within which the soil name occurs. Appropriate definitions for the possible values may be found in "A Proposed Soil Mapping System for Canada".

SOIL CLASSIFICATION OF DOMINANT PEDON - is recorded at the subgroup level in accord with the "System of Soil Classification for Canada".

PHASES - up to four phases may be defined by circling the appropriate self-defining codes. This is the only variable (data field) for which more than one value can be recorded.

For the remaining data to be recorded about a soil name in the fields, the definitions and conventions have been documented in the "Canada Soil Information System (CanSIS) Manual for Describing Soils in the Field (revised 1982). The fields are as follows:

PARENT MATERIAL I, II & III (9 fields);

PHYSICAL COMPONENT
CHEMICAL COMPONENT
MODE OF DEPOSITION
LANDFORM CLASSIFICATION SURFACE EXPRESSION (1 field)

Table 3: CanSIS Codes for Canadian Provinces & Territories

Province	Input Code
British Columbia	01
Alberta	02
Saskatchewan	03
Manitoba	04
Ontario	05
Quebec	06
Newfoundland	07
New Brunswick	08
Nova Scotia	09
Prince Edward Island	10
Yukon Territory	11
Northwest Territories	12

MINERAL SOIL FAMILY CRITERIA FOR DOMINANT PEDON (7 fields);

PARTICLE SIZE OF CONTROL SECTION
MINERALOGY OF CONTROL SECTION
SOIL DEPTH
REACTION OF C
CALCAREOUSNESS OF C
SOIL TEMPERATURE
SOIL MOISTURE

ORGANIC SOIL FAMILY CRITERIA FOR THE DOMINANT PEDON (8 fields);

ORGANIC SURFACE TIER
MINERAL SURFACE TIER
REACTION OF CONTROL SECTION
SOIL TEMPERATURE
SOIL MOISTURE
PARENT SIZE OF TERRIC LAYER
LIMNIC MATERIALS
DEPTH TO LITHIC CONTACT

SUBCLIMATE - The subclimate represents a definition in 4 characters of the specific climatic features to be associated with a soil name. There are no nationally defined classes or conventions for the use of this field; however, its use should be coordinated at least at the provincial level.

When values for all relevant characteristics to be associated with a soil name have been recorded the data are verified within the province and then forwarded to the national soil correlator for input to the national soil names file. Upon receipt of data approved by the national correlator, a data coordinator will proceed to input the data into the computer as outlined in Table 2.

6. OUTPUT AND RETRIEVAL OF DATA FROM THE CANSIS CUSTOM FILE MANAGEMENT SYSTEM

6.1 Overview

As CanSIS started, a generalized file management system was custom developed to accommodate the needs of the Land Resource Research Institute in maintaining an inventory of soil information from across Canada. It allowed the user to define one or more data files, to perform the transactions that will create and update information in the files, and to retrieve information from the files in both an ad hoc informal method and in more formatted output such as reports.

Each data file can be accessed individually for data retrieval. Any or all of its fields may be used for selection or output and any of its numeric fields can be used for in any mathematical calculation. In addition, subfiles can be defined and created and formal outputs such as reports and profiles can be defined and executed. Reports allow for automatic summarizing and calculating at specified breakpoints as well as for fully formatted output.

There is no single generalized set of procedures for data retrieval from the file management system (FMS).

Consequently, a major portion of the FMS is the Data Retrieval Language (DRL) which is used by the CanSIS programming staff to effect retrieval requests. This language allows the programmer to retrieve data, produce a file and then hand the data to a Report Generator, RAPID, EASYTRIEVE, SAS or other commercial packages.

Each information request must be defined using the DRL and entered into the CanSIS data retrieval program.

Requests will produce a printed listing of records, segments and fields from the CanSIS FMS; CanSIS type subfiles of any CanSIS data file; and flat files made up of single values from any fields in the CanSIS data base. The flat files can be passed to the CanSIS report generation program or to other packages.

It is beyond the scope of this manual to provide full documentation on DRL applied to the file management system. A description of the DRL command statements is contained in the FMS manual.

6.2 Profile Description Routine

A major program in the Detail file of CanSIS is the Profile Description definition program.

This program was designed specifically to accommodate a conventional form used by many pedologists to describe an individual soil site and profile. This output consists of a series of paragraphs describing the general site characteristics, secondly, a statement for each soil horizon at the site, and thirdly, a series of tables which contain analytical laboratory data by horizon.

Other reports can be developed. The amount of time required to develop these depends on the complexity of the output required. Since the program is also responsible for grammar and punctuation, instructions are very precise and account for all data situations.

Details in developing other reports are contained in the CanSIS File management System Users Guide. Details on instituting a Profile description run are contained in appendix 9 of this manual.

6.3 Flat File Definition

The flat file represents a generalized set of procedures for accessing data in CanSIS FMS files. It uses the data retrieval language associated with the FMS to retrieve data from specific fields and segments and places it in a sequential file in standard format. From this sequential file or "flat" file the data can be read or defined to other programs for further computer manipulation and report preparation. This procedure is normally only carried out by the CanSIS programmer/analyst staff in Ottawa. It is briefly described in Appendix 10.

7. RAPID DATA BASE MANAGEMENT SYSTEM

7.1 Overview

RAPID is a generalized data base management system (DBMS) developed within Statistics Canada as a common solution to many statistical data processing applications. The following objectives are selected from the RAPID documentation:

1) To promote a consistent approach to data storage across application boundaries as a prelude to large scale data integration.
2) To encourage data structuring that is both simple to understand for the non-systems expert, and effective for the increasingly complex requirements placed on it.
3) To allow systems to evolve, as requirements change, without the loss of control over existing data capital and expensive system re-development.
4) To provide a common interface for both generalized and custom programmed application systems.
5) To provide a completely general data base management system with integrated data dictionary facilities and utilities to cover the common needs of application systems.

Within CanSIS, the objectives to be met with the output format of data organization were to allow for (i) complete reports of individual sites, experiments or studies (ii) summaries and tabulations of results from a selection of data, (iii) easy interface of the data into commercial statistical software packages (iv) spatial representation of data with a geographical reference in a form suitable for overlaying onto a soils map or other coverage, (v) formatting the data into a machine readable form so that it could be transferred easily onto other computer systems for additional manipulation, (vi) general user access to the data in a form in which it could be sorted, selected and manipulated. Of these objectives, one of the major considerations was the need to facilitate ON-LINE ITERATIVE MANIPULATION AND SELECTION OF THE DATA BY THE GENERAL USER:

The easiest way to ensure that this final objective was achieved was to organize the data into a series of relations in a data base management system (DBMS) called RAPID.

All RAPID relations in CanSIS are identified by a unique four character name, as indicated in Table 4.

Table 4: GENERAL RELATION NAME CONVENTIONS IN CANSIS

Computer Legends	E040 E100 etc.	The 'E' indicates a computerized Extended Legend and the number designates a specific legend.
Soil Data Relations (Detail)	SD01 SD02 etc.	The 'SD' represents Soil data and the number distinguishes one from the other.
Performance Management	PMA1 PMA2 etc.	The 'PM' represents Performance Management relations and the A1 to I4 distinguish different relations.
Soil Names	SL01	This designates the single Soil Names relation.
Project and Daily Relations	MN01 SK01 etc.	The first two characters, usually letters, designate a province; the number separates numbers of a series.
Parks Canada Relations	W000 W001 etc.	The 'W' represents observations by Park Wardens.

The conventions for assigning these codes are not absolutely rigorous, but generally the first two characters identify the information type and the last two the order of the information.

7.1.1 RAPID Relation Structure - An Example from the Performance/Management File.

As defined in RAPID, a relation consists of a table of data with two major axes; one a key or identification axis and the other a characteristic axis containing all the properties or attributes which are associated with the key. A part of the first relation of the P/M file is illustrated in the following example:

Example 1: Relation pmal

Key					CHARACTERISTICS OR ATTRIBUTES			
Province	Expyear	Projid	Agency	Suprvisr	Expstart	Expend	Exppurp	Slpprcnt
01	68	12345	3	AB	66	69	Fert-resp	4
02	71	1BB01	2	AH	70	73	Fert-resp	

etc.

Each observation must contain a unique key. However, fields within a key may be used for retrievals.

Each field name represents an abbreviation of the type of information contained in that column of the relation. It represents the name which must be specified to the computer to access data contained in that column. In Example 1 samples of the type of information which might be output are illustrated. For the key fields the information output is identical to the input. i.e.

Field name

PROVINCE	indicates the province	01 represents British Columbia 02 represents Alberta etc.
EXPYEAR	represents the year of the experiment	

PROJID	a five character codedesignating the experiment
AGENCY	indicates the agency '3' represents university conducting the experiment
SUPRVISR	gives initials of the supervisor of the study.

The purpose of the key fields is to define precisely and uniquely each record in the file. Information in the attribute or characteristic fields can take a variety of forms:

EXPSTART	is the year the experiment started
EXPEND	is the year of experiment termination
EXPPURP	explains the purpose of the experiment. This information is input as the code '2'. It is stored as an internal code assigned by the RAPID system, and output as the connotativelabel "Fert-resp" to represent a fertilizer response trial.
SLPPRCNT	indicates the percent slope. The value in this field must be a number.

7.2 Types of Fields and Implications

Information recorded in a RAPID relation may be described by four types of fields; character, numeric, coded, logical. The choice of field type must be made at the time data are organized into relations. The implications regarding the choice are presented below.

7.2.1 Character fields have no restrictions on the values which they contain (provided the characters exist on the input keyboard). The range in values is infinite and may include blank or missing values. Although there are no restrictions on input, there are limits to the kinds of operations which may be performed with them. Information in character fields can be displayed, or used for selection by specifying individual values or values greater than, less than or within a range. Character fields can be sorted or ordered alphanumerically.

7.2.2 Numeric fields are restricted to numbers, either with or without a sign. Values may include the infinite range of numbers but blanks are not allowed and a specific value must be assigned if data are missing. In addition to operations which can be performed with character fields, numeric fields can be specified for arithmetic operations e.g. sum, means etc.

7.2.3 Coded fields have a limited, predefined range of permitted values. Arithmetic operations cannot be performed on these fields. However, coded fields constitute a classification and can be used to group other data into classes as defined by the coded values.

When selection criteria are addressed to a coded field they must specify the specific coded values to be selected; ranges, greater than or less than conditions are not acceptable.

7.2.4 Logical fields are fields which have two possible values; a "1" or a "0" i.e. a "yes" or a "no". The value in these fields can either be displayed or used as part of a selection.

In theory, it is possible to represent any type of information in one or a series of relations using combinations of the above kinds of fields.

7.3 RAPID Data Dictionaries - Their Use and Interpretation

7.3.1 Overview

One of the steps in representing data in one or a series of RAPID relations is the creation of a RAPID data dictionary. While the principal function of the data dictionary is to define to the computer the format and characteristics of the data, the computer listing resulting from dictionary creation also serves as a guide to the user for data retrieval.

An understanding of the contents and organization of a data dictionary is a necessary step towards accessing and using the data in a RAPID relation. In the following sections a computer generated data dictionary is described and interpreted. Examples are taken from the Soil Names File of CanSIS; however, the principles of using a standard computer-generated data dictionary are common to all files.

A RAPID data dictionary contains the following sections;

(i) Table of contents - a listing of the contents of the dictionary and their page reference (illustrated in figure 2). There is no fixed sequence for the major sections (codeset, valueset, relation and file); listings under these sections occur in alphabetic order.
(ii) RAPID files - definitions of the standard sequential computer files which specify the format and characteristics of the input data.
(iii) Codesets - define the codes used in the input file and give the corresponding output label.
(iv) Relations - a listing for each relation in the dictionary of field names, type and a descriptive comment.
(v) Valuesets - a listing of the labels associated with coded fields and possibly a descriptive comment.

In CanSIS, most data dictionaries contain definitions for a series of relations, their related files (one or more per relation), codesets and valuesets. Relations from a single national file, or for a map series are commonly grouped into a dictionary. Definition of the dictionary is the first step in creating a relation. Subsequently, copies of the dictionary can be produced as required and used as a reference document.

Figure 2: Table of contents (Soil Names File Data Dictionary Version June'83).

Only current versions of data dictionaries should be used as reference documents for data input, manipulation or retrieval from RAPID relations. Data dictionaries can be obtained on request from the CanSIS Project Leader who will either provide an up-to-date dictionary or the computer program necessary to route a dictionary listing directly to the user.

7.3.2 Data Input and Transfer

For users who wish to input data to existing relations or who require the data in a known and documented format for transfer to other installations, the definitions of the RAPID files and codesets provide the information required.

7.3.2.1 RAPID File: A file is made up of a number of records. In the RAPID file definition sections of the data dictionary, the format of the individual records, are described column by column. This description refers to the machine readable form of the data normally entered on computer input forms. The data organization on the RAPID file record will not necessarily resemble the original input document layout. This is particularly true for data which are initially input to the CanSIS file management system which generally require extensive reformatting for input to RAPID.

The definitions of RAPID files in the data dictionary contain five components of interest to the user. Figure 3 illustrates a file definition for the soil names file (as it existed in June 1983) and identifies the five parts as follows.

A. 'SL01A' designates the file name. Normally this name will be five characters in length and the first four will be identical to the corresponding relation.

B. Field Name. The field name is a descriptive or characterizing name up to 8 characters long used to identify the contents of one or a group of columns on the input record (a field or variable) to the user. e.g. CALCC identifies the name given to data describing the calcareousness of the C horizon.

C. Position. The column position of each field on the input record is designated as POS(XX) where XX is the column in which data for this field starts.

D. Type and Length. In the majority of RAPID dictionaries a field will be defined as character, text, fixed binary, or fixed decimal. The abreviations used for these are CHAR, TEXT, FIXED BIN, or FIXED DEC respectively. The number following the designation indicates the length of the field in columns (bytes actually) for character or text fields; as binary positions for fixed binary fields and as positions and number of decimal places for fixed decimal fields (e.g. 8,3).

E. Codeset Names. This section contains information only when the field is coded and has an associated codeset containing the limited range of values possible in this field. Entries in this section have the form CODED (CZZZZZ) where CZZZZZ is the name of a codeset listed in the table of contents.

Figure 3: RAPID file SL01A (from Soil Names File Data Dictionary version June '83).

A			
FILE	SL01A		
B	C	D	E
CALCC	POS(69)	CHAR(1)	CODED(CCALCC)
CLASSDP	POS(38)	CHAR(3)	CODED(CCLASSDP)
DPTHLTHC	POS(77)	CHAR(1)	CODED(CDPTHLTHC)
LEVEL	POS(37)	CHAR(1)	CODED(CLEVEL)
LIMNMAT	POS(76)	CHAR(1)	CODED(CLIMNMAT)
MINCS	POS(65)	CHAR(2)	CODED(CMINCS)
MINSURTR	POS(73)	CHAR(1)	CODED(CMINSURTR)
ORGSURTR	POS(72)	CHAR(1)	CODED(CORGSURTR)
PARSZCS	POS(62)	CHAR(3)	CODED(CPARSZCS)
PARSZTL	POS(75)	CHAR(1)	CODED(CPARSZTL)
PHASE1	POS(41)	CHAR(1)	CODED(CPHASES)
PHASE2	POS(42)	CHAR(1)	CODED(CPHASES)
PHASE3	POS(43)	CHAR(1)	CODED(CPHASES)
PHASE4	POS(44)	CHAR(1)	CODED(CPHASES)
PM1CHECO	POS(47)	CHAR(1)	CODED(CPMCHECO)
PM1MDEP	POS(48)	CHAR(2)	CODED(CPMMDEP)
PM1PHYCO	POS(45)	CHAR(2)	CODED(CPMPHYCO)
PM2CHECO	POS(52)	CHAR(1)	CODED(CPMCHECO)
PM2MDEP	POS(53)	CHAR(2)	CODED(CPMMDEP)
PM2PHYCO	POS(50)	CHAR(2)	CODED(CPMPHYCO)
PM3CHECO	POS(57)	CHAR(1)	CODED(PMCHECO)
PM3MDEP	POS(58)	CHAR(2)	CODED(CPMMDEP)
PM3PHYCO	POS(55)	CHAR(2)	CODED(CPMPHYCO)
PROV	POS(7)	CHAR(2)	CODED(CPROV)
PROVCD	POS(7)	CHAR(2)	
PROXUSA	POS(35)	CHAR(1)	CODED(CPROXUSA)
REACC	POS(68)	CHAR(1)	CODED(CREACC)
REACCONS	POS(74)	CHAR(1)	CODED(CREACCONS)
REPORTNO	POS(32)	CHAR(3)	
SHRTNAME	POS(12)	CHAR(16)	
SOILCODE	POS(9)	CHAR(3)	
SOILDPTH	POS(67)	CHAR(1)	CODED(CSOILDPTH)
SOILMOIS	POS(71)	CHAR(1)	CODED(CSOILMOIS)
SOILNAME	POS(12)	CHAR(20)	
SOILTEMP	POS(70)	CHAR(1)	CODED(CSOILTEMP)
SOLSBGRP	POS(38)	CHAR(3)	CODED(CSOLSBGRP)
STATUS	POS(36)	CHAR(1)	CODED(CSTATUS)
SUBCLIM	POS(78)	CHAR(4)	
SURFEXP	POS(60)	CHAR(2)	CODED(CSURFEXP)

7.3.2.2 Codesets: The codeset is the table of all possible values a coded field can have. It is made up of four parts as illustrated (figure 4) in the codeset CPROV from the Soil Names File:

A. Codeset name CPROV. The name of each codeset always starts with the letter 'C' and is referred to in the RAPID file, the table of contents and the codesets.

B. Input Code. The input code specifies the acceptable input data for this field.

C. Output Label. The output label lists the label or name corresponding to each input code. This label is passed on to the relation for output. It is restricted to 16 characters in length and cannot contain any imbedded blanks.

D. Descriptive Comment. In some codesets there will be a descriptive or explanatory comment to the right of the input code. It is optional and not present in the example.

7.3.3 Data Retrieval and Output

For users wishing to access data in a RAPID relation, the data dictionary provides a list of field names, their type and a descriptive comment. The field names must be used for retrieval. In addition for the coded field, the labels or values are listed. If a selection is to be based on values in a coded field, these lists represent the only acceptable values. The sections of a data dictionary which pertain to data retrieval are the relation definitions and the valueset listings.

A convention adopted by CanSIS is that fields with the suffix 'CD' designate fields in which the input code is stored. They contain the same information as fields without the suffix, e.g. a record for which the value of the field PROV is Quebec would have a value of '06' in the field PROVCD. The data are in fact listed twice from the one input record in this case '06'. This redundancy is convenient for some types of retrievals.

All relations have a special field '$INDEX' which does not appear in the data dictionary. $INDEX is a unique identifier like the key because it refers to the internal row number of data in the relation. It is particularly useful to restrict scanning and selection type retrievals from relations with many lines of data.

7.3.3.1 Relations: Each relation description is made up of four parts as illustrated in figure 5.

A. Relation Name 'SL01'. All RAPID relations are denoted by a four character name. It is this name which is used to designate to the computer which relation is required.

B. Field Name is the name by which attributes or properties are specified.

Figure 4: CODESET CPROV (from Soil Names File Data Dictionary version June '83).

A		
CODESET	CPROV	
	B	C
	ALBERTA	'02'
	BLANK	' '
	BRITISH-COLUNBIA	'01'
	MANITOBA	'04'
	N.W.T.	'12'
	NEW-BRUNSWICK	'08'
	NEWFOUNDLAND	'07'
	NOVA-SCOTIA	'09'
	ONTARIO	'05'
	P.E.I.	'10'
	QUEBEC	'06'
	SASKATCHEWAN	'03'
	YUKON	'11'

Fisure 5: Relation S101 (from Soil Names File Data Dictionary version June '83).

A

RELATION SL01

B	C	D
CALCC	CODED(VCALCC)	CALCAREOUSNESS OF C
CLASSDP	CODED(VCLASSDP)	SOIL CLASSIFICATION OF DOMINANT
DPTHLTHC	CODED(VDPTHLTHC)	DEPTH TO LITHIC CONTACT
LEVEL	CODED(VLEVEL)	LEVEL
LIMNMAT	CODED(VLIMNMAT)	LIMNIC MATERIAL
MINCS	CODED(VMINCS)	MINERALOGY OF CONTROL SECTION
MINSURTR	CODED(VMINSURTR)	MINERAL SURFACE TIER
ORGSURTR	CODED(VORGSURTR)	ORGANIC SURFACE TIER
PARSZCS	CODED(VPARSZCS)	PARTICLE SIZE OF CONTROL SECTIO
PARSZTL	CODED(VPARSZTL)	PARTICLE SIZE TERRIC LAYER
PHASE1	CODED(VPHASES)	PHASE-CARBONATED
PHASE2	CODED(VPHASES)	PHASE-LITHIC
PHASE3	CODED(VPHASES)	PHASE-PEATY
PHASE4	CODED(VPHASES)	PHASE-SALINE
PM1CHECO	CODED(VPMCHECO)	PARENT MATERIAL 1 - CHEMICAL CO
PM1MDEP	CODED(VPMMDEP)	PARENT MATERIAL 1 - MODE OF DEP
PM1PHYCO	CODED(VPMPHYCO)	PARENT MATERIAL 1 - PHYSICAL CO
PM2CHECO	CODED(VPMCHECO)	PARENT MATERIAL 2 - CHEMICAL CO
PM2MDEP	CODED(VPMMEDP)	PARENT MATERIAL 2 - MODE OF DEP
PM2PHYCO	CODED(VPMPHYCO)	PARENT MATERIAL 2 - PHYSICAL CO
PM3CHECO	CODED(VPMCHECO)	PARENT MATERIAL 3 - CHEMICAL CO
PM3MDEP	CODED(VPMMDEP)	PARENT MATERIAL 3 - MODE OF DEP
PM3PHYCO	CODED(VPMPHYCO)	PARENT MATERIAL 3 - PHYSICAL CO
PROV	CODED(VPROV)	PROVINCE
PROVCD	CHAR(2)	PROVINCE
PROXUSA	CODED(VPROXUSA)	PROXIMITY TO U.S.A.
REACC	CODED(VREACC)	REACTION TO C
REACCONS	CODED(VREACCONS)	REACTION TO CONTROL SECTION
REPORTNO	CHAR(3)	REPORT NUMBER
SHRTNAME	CHAR(16)	SHORT NAME
SOILCODE	CHAR(3)	SOIL CODE
SOILDPTH	CODED(VSOILDPTH)	SOIL DEPTH
SOILMOIS	CODED(VSOILMOIS)	SOIL MOISTURE
SOILNAME	CHAR(20)	SOIL NAME
SOILTEMP	CODED(VSOILTEMP)	SOIL TEMPERATURE
SOLSBGRP	CODED(VSOLSBGRP)	SOIL SUBGROUP
STATUS	CODED(VSTATUS)	STATUS
SUBCLIM	CHAR(4)	SUBCLIMATE
SURFEXP	CODED(VSURFEXP)	SURFACE EXPRESSION

C. This portion of the relation designates either
   (i) that a field is coded and the possible range of values are listed in a valueset designated in parenthesis (C1) or
   (ii) the type and length of a field (C2)

D. Descriptive Comment. This section may or may not be present. When present it attempts to clarify the nature of the field or in the case of numeric fields list the value representing missing values.

7.3.3.2 Valuesets: The valueset is a list of valid output labels for coded fields (figure 6). It is made up of three parts.

A. Valueset Name 'VLIMNMAT'. The name of each valueset always starts with the letter 'V' and is referred to in the table of contents, the relation definition and the valuesets.

B. Value. The value or label used for retrievals. The values are restricted to 16 characters and cannot contain imbedded blanks. NOTE: In addition to the values listed in the valuesets, data retrieval from coded fields may contain the value 'NULL'. This represents input data for which the input code was invalid and as a consequence no valid output label exists.

C. Descriptive Comment. This comment may or may not be present and serves to clarify the label.

Figure 6: Valueset VLIMNMAT (from Soil Names File Data Dictionary version June '83).

	A	
VALUESET	VLIMNMAT	
	B	C
	BLANK	NO DATA
	COPROGENOUS	COPROGENOUS
	DIATOMACEOUS	DIATOMACEOUS
	MARL	MARL

8. RAPID RELATION OUTPUT AND RETRIEVAL, USING EXAMPLES FROM THE SOIL NAMES FILE

With the exception of the cartographic file, all national files, many project or daily files and all computer legend files of CanSIS have been defined into the RAPID data base management system and organized into one or a series of relations. In this section the procedures for data retrieval and output are described and illustrated with examples from the Soil Names File.

In addition to the documentation contained in this section, users attempting to access and retrieve data from the Soil Names File should:

(i) review section 5.3.1 contents of the Soil Names File and also the computer input form Appendix 3.
(ii) obtain from CanSIS, Ottawa the most recent copy of the Soil Names File Data Dictionary (if it was not received as a supplement to this manual).
(iii) review the previous section describing the use of a computer generated data dictionary for data retrieval and output.

8.1 Soil Names File Data Organization

The Soil Names File of CanSIS catalogues all soil names (past and present) and their principal attributes. In this way it serves as a guide to information contained in CanSIS on a particular soil.

The Soil Names File is defined on RAPID as a single relation with the four character code SL01.

The structure of the relation consists of a table of data where variables or characteristics form the vertical columns and observations form the horizontal rows. In this file the data records are identified by their province and soil series code; these are the 'key' fields and must be unique for each record.

8.2 Data Retrieval

This section of the manual is designed as a reference for those who wish to output information from RAPID relations. It contains a detailed description of the procedures which allow the user to select and display the information at a terminal. In associated appendices are short explanations of standard types of reports, that have been established using EASYTRIEVE (a commercial software package). A detailed discussion of the use of commercial packages is beyond the scope of this manual and the user is referred to the adequate commercial documentation supplied with them.

NOTE: All computer-generated statements appear in upper case letters, and all statements entered at the terminal appear in lower case letters and are followed by pressing the RETURN key. The "?:" is computer generated and is called a DREAM prompt. Comments from the authors of this manual are in parenthesis.

8.2.1 DREAM (Basic Level Output)

8.2.1.1 Set-up for retrievals for DREAM: A variety of generalized computer program packages can be used to access a RAPID relation e.g. the Soil Names File. The simplest and most widely used package is called DREAM (Direct Relational Access Method), which constitutes part of the RAPID software obtained from Statistics Canada. It provides on-line, cost effective, quick access to data which can be easily searched to determine whether it is of interest and warrants further manipulation.

In DREAM, a user can specify selection criteria and examine a subset of information. He can count or display any field in the relation SL01 (Soil Names).

Establish contact with Datacrown (Appendix 2)

```
READY
ex sis01(lsisdrm)          (connects the user to DREAM in CanSIS)

ARE YOU A NEW USER?
no                         If no is specified, the computer moves
                           directly to DEFAULT PAGE WIDTH)
                           see appendix 2 for listing

DEFAULT PAGE WIDTH = 79
DEFAULT PAGE DEPTH = 20
WOULD YOU LIKE TO ALTER THESE?

(Unless these defaults are changed, the largest block of data the
computer will print at one time will be 79 characters wide and 20
rows deep. This is suitable for a video terminal. Hard copy
terminals may use a width of up to 129 characters)

yes   (if no, the computer will move directly to FILES AVAILABLE)

ENTER PAGE WIDTH:
129

ENTER PAGE DEPTH:
100  (The maximum number of rows printed at one time)

RELATIONS AVAILABLE (DATE):  CP01, CP02, CP03, CP04, CP04, CP05,
                             E027, E028, E029, E030, E040, SOIL,
                             PME1, PMF1, PMG1, PMI1, PMI3, PMI4,
                       ALSO: E005, E006, E007, E008, E009, E010,
                             W001, W002, W003, W004, SL01, SD01,
                             SD02, SD03

FILE(S) FREED:  SYS00001   SYS00009
WOULD YOU LIKE TO ALLOCATE A REQUEST LIBRARY?
no
```

```
RELATIONS AVAILABLE (DATE):  CP01, CP02, CP03, CP04, CP05, E027,
                             E028, E029, E030, E040, SOIL, PME1,
                             PMF1, PMG1, PMI1, PMI3, PMI4,
                      ALSO:  E005, E006, E007, E008, E009, E010,
                             W001, W002, W003, W004, SL01, SD01,
                             SD02, SD03

FILE(S) FREED: SYS00001   SYS00009
WOULD YOU LIKE TO ALLOCATE A REQUEST LIBRARY?
no
```

(This capability is not necessary for the present purpose of familiarization with DREAM and with this file)

8.2.1.2 Dream Commands: Entering a question mark after the Dream prompt sign '?:' will produce a list of all the fields that are present in the relation specified.

It is normally desirable to place limiting parameters on the data to be retrieved. This is achieved with DREAM's Boolean facilities by using a value operand along with a primary command. Several delimiters may be used in one request, if desired, but each selection statement should be entered as a separate line. Without limiting the selection or setting criteria, all records (observations) in the field will be displayed. The following list of operands can be used in either symbol or letter form:

= eq (denotes equal to; use with coded, character or numeric fields). A maximum of 9 fields can be specified with this operand.
ne (denotes not equal to; use with coded, character or numeric fields)
> gt (denotes greater than; use with numeric or character fields
< = GE (denotes greater than or equal to; use with numeric or character fields)
< lt (denotes less than; use with numeric or character fields)
< = le (denotes less than or equal; use with numeric or character fields)
>< in (denotes within range; use with numeric or character fields)
<> ex (denotes outside range; use with numeric or character fields)

For example, prov = ontario will select the data that has the value ontario under the field prov. Several values may be requested after an operand interpreted as "or". These are separated by commas, e.g. prov = ontario,quebec. Two operands must be entered on separate lines. The computer will interpret these as "and". (See example 3, Appendix 5).

Two primary commands are used for retrieving information in the relation SL01 (Soil Names): DISPLAY and COUNT: They can be used alone or together, but the display will always be printed first.

When using the display command the user must specify which field or fields he wants printed. The computer will list in tabular form the data in those fields. (See example 1 appendix 5).

The count command will give a total of the number of records in the file. It can be used for retrieving data for any type of field. Also it can show the distribution of data in coded fields. (e.g. entering "count soil name by surfexp" would produce a distribution table with the values of surfexp displayed with the number of soil names beside each value. See example 2, appendix 5.

If the computer is asked to print out a longer line of data than can be displayed by the designated page width during data retrieval, it will give the name of the first "over hanging" field and say:

(field name) DROPPED, LINE EXCEEDS MAXIMUM LENGTH

This means that data for the fields before the named field will be printed. Data for the other fields may be obtained by entering them as a separate request.

When all commands for one group of data have been entered, type 'go' after the computer's ?: to initiate the retrieval. The method of retrieval will be displayed and data output will follow.

At the end of every 20 lines of data printout (more or fewer lines if default page depth was changed) the computer will print 'END' OR 'RETURN'. If the Return key is depressed, the output of data will continue. If end is entered, no further output for that request will be printed.

When all the data in a request have been printed, the computer will then print:

(RETRIEVAL COMPLETE)
ENTER 'END' OR NEXT REQUEST
?:

A new request may now be entered. It should be noted that the computer does not retain any of the limits or commands from a previous request.

8.2.1.3 Mistakes and Error Messages: Mistakes are common during the course of a data request. If a mistake is discovered before the Return key is depressed, it is possible to backspace to the error and enter the proper character. However, each character that is passed during backspacing will be erased. If the error is discovered after the Return key has been depressed, then depress the BREAK key followed by the RETURN key. The computer will now respond with:

ENTER SKIP, CONTINUE OR HALT:

SKIP - If SKIP or S is entered or if the Return key is depressed, the preceeding request will be cancelled and the computer will be ready for a new request.

CONTINUE - If CONTINUE or C is entered, the computer will continue to process the data from the point at which the BREAK occurred. However, some data output may be lost.

HALT - If HALT or H is entered, the computer will move out of the soil file, but will still be in DREAM, and will respond with:

DO YOU WISH TO ACCESS ANY OTHER FILE?

The Soil Names Relation may be re-accessed by responding with yes, and entering SL01 after the appropriate generated prompt.

COLUMN NOT ON RELATION

This response after a command indicates that a variable or field has been entered which is not recognized by the computer. This is often caused by spelling or context (spacing) errors on input.

INVALID OPERATOR

This means that an invalid command has been given to the computer; check spelling and context.

CNF
Again, check for spelling errors.

DUPR RETURNED FROM $DEFR

This often occurs after a BREAK-SKIP sequence. 'Go' should be entered, after which the computer will respond with:

REQUEST CONTAINS NO OUTPUT STATEMENT
(RETRIEVAL BEGINS:)
(RETRIEVAL COMPLETE)
ENTER 'END' OR NEXT REQUEST
?:
The request may now be re-entered.

8.2.1.4 Exiting from DREAM: When you have finished doing your retrievals and you receive the message:

(RETRIEVAL COMPLETE)
ENTER 'END' OR NEXT REQUEST
?: end (This instructs the computer to move out of the Soil Names Relation)

DO YOU WISH TO ACCESS ANY OTHER FILE?

no (This instructs the computer to move out of DREAM).

READY (The computer is now in the READY mode at the Datacrown facilities).

logoff (This logs the computer off)

AB123CD LOGGED OFF AT 14:22 ON FEB 15, 1983 (The computer is now no longer in contact with the terminal).

NOTE: If using DATAPAC make sure to disconnect the call.

8.2.2 Output using Standard Software Packages

DREAM constitutes the only output software available to the user as part of RAPID. Capabilities are developed to output data using other report writing and statistical packages such as EASYTRIEVE, SAS, SPSS, etc.

Use of packages such as these require a greater input on the part of the user to develop more computer interaction skills either on his own or with the assistance of a programmer/analyst.

It is not the purpose of this manual to define the wide range of data manipulations nor to describe the possible programming steps necessary to carry out all possible types of analysis required for data in the various CanSIS files. In general data manipulations more complicated than those possible with DREAM software will be carried out with commercially available software packages or an appropriate compatible standard programming language. This latter will be the case only if the user is knowledgeable in the language. Where commercial software packages are used the associated commercial documentation should provide adequate direction on the types of reporting and data analysis possible. Within CanSIS two software packages have been used extensively.

(i) EASYTRIEVE - primarily for simple data reorganization and report writing.
(ii) SAS (Statistical Analysis System) for full statistical analysis and graphical representation.

In contrast to DREAM [in which the data manipulation commands are applied directly to the data by the statement "go" and for which the results of the analyses are written immediately (i.e. on-line data interaction)] both EASYTRIEVE and SAS are used to manipulate data in CanSIS in "batch mode".

Batch mode operation means that the entire request is defined (including specification of the RAPID relation or data file, the data to be selected and the manipulations desired; the output report and its format). The entire request is then submitted as a unit for processing. Normally the output of a batch job is printed at a high speed printer but it can be routed for access at a terminal.

The types of access and reports available through use of general software packages have been documented. Some sample procedures and reports are documented in Appendix 6. The programs are available to researchers to output standard reports and to provide examples of more specialized reports.

8.2.3 Spatial Display

All records in CanSIS which relate to sites (i.e. all files except soil names, cartographic, and computer legends) contain a location description. Generally, this location takes the form of geographic coordinates in either the universal transverse-mercator designation or in degrees latitude and longitude. Either of these formats is

appropriate for computer manipulation and are the only ones acceptable for spatial display. The spatial display point plot package (AGSlSPPL) is a generalized set of programs and procedures for plotting data points and, if desired, appropriate labels within a defined boundary. Its principal application is in plotting data from CanSIS files at locations which are correct relative to some available map scale and projection. The plot can then be overlayed and registered to the appropriate map to give a spatial display of tabular information.

The actual plot preparation is normally carried out by the CanSIS staff in Ottawa based on information provided by the person requesting the plot. A sample of the forms specifying the information required is listed in Appendix 7. Prior to completing this form the information user will normally access the data to be plotted using the software package DREAM. In this way he can confirm that the number of data points are adequate and determine whether the data should be grouped into classes for plotting.

A single plot can contain a variety of types of data.

The size, colour and type of points are variable. Up to 14 different shapes and symbols can be used to reference different types of information. Labelling of these points is optional.

Information may be presented exactly as it exists in the data file, i.e.; labels may be actual numbers or names of observations, or they may be grouped into classes and plotted. Colours may be used to identify the group of symbols for example: red may be used for both x and + and green for both y and z; or, to identify classes within a symbol for example a red + may represent one to five observations and a blue + may represent 6 to 15 observations.

Note that the points are always plotted directly at the locations; there is no provision to offset different groups (e.g. if red represents female and blue represents male they would be superimposed and indistinguishable in the plot).

If desired, a "label" can be associated with each observation, up to 30 characters in length and one space to the right of the symbol.

A legend and title are specified by the user. Three lines are allocated for the title. Size of letters is usually .25 inches and location is usually 1 inch below left hand corner of the plot.

Up to 15 lines of legend can be specified. Size of letters are usually the same as labels, i.e. .10 inches.

Standard information that need not be included in title or legend (because it is plotted automatically)

a) date and time
b) boundary map ID
c) plot ID
d) scale
e) number of points plotted

Virtually any scale can be selected. Realize the overall dimensions of the plot.

These spatial display procedures can be used to plot any type of data which has an acceptable location code. Complete technical details for operating the procedures are documented in appendix 8. (Note this appendix exists in mimeographed form only and should be requested as item 8 Appendix 1.

9. FLOW OF DATA INTO THE CARTOGRAPHIC FILE

9.1 Cartographic Data Input Formats and Conventions

The process of inputting soil maps into the Cartographic subsystem begins at the initial stages of a mapping project. The Cartographic unit of LRRI have documented procedures which will ensure that the map products produced for computer input are of high quality. Adherence to procedures will also ensure that the progress of the map through the various stages will continue uninterrupted and require minimal effort in the input and error correction stages. The originator will have access to data, acreages and interpretive maps at the earliest date possible.

Any thematic information displayed in map form can be input, stored, manipulated and retrieved from the CanSIS cartographic file. However, there are some restrictions on the size of maps and the level of detail which can be input and in particular some standards of base map quality and manuscript preparation which must be met. Manuscript maps are normally submitted by regional soil survey units and both the map and legend are subject to correlation to ensure that the symbolization and format correspond to nationally acceptable standards. These teps are required because the cartographic file is a national file. A different approach could be taken for special project maps. The normal steps in the computerization of cartographic data are outlined in Table 5.

The first step in the procedures is ordering a proper base map on which to represent the thematic data. All the necessary details (specifications, forms, etc.) of ordering a base map are contained in the publication "Base Map Requisition/Demande de Fond de Carte" available from the Cartography unit of LRRI.

The unit has access to all base material through Energy Mines and Resources and provincial agencies. Non-standard scale bases can be produced in-house. Since base map preparation is time consuming, it is recommended that material be ordered early in the planning stages of the survey.

If provincial bases are preferred, printing negatives should be borrowed on a short term loan, and sent to the Cartography section of LRRI. Alternatively if the name of the provincial authority who requested the map can be supplied, the Cartographic unit will obtain the necessary materials.

Table 5 Flow of Information into the Cartographic File.

1) Submission of manuscript map in final form to the cartography section. This manuscript must conform to the standards set out by the Cartographic Section including guidelines on acceptable base maps, etc. These standards should be on hand prior to the map compilation.

2) Approval of the manuscript and assignment of priority.

3) Digitization of the thematic information and input of the map symbols. The input of cartographic information requires special hardware, software and personnel. Manuscript maps are currently input to CanSIS only at installations in Ottawa and Manitoba. In all cases the final input of data to the national cartographic file is carried out in Ottawa.

4) Computer check for errors.

5) Error correction.

6) Manual edit for errors by CanSIS cartographic staff.

7) Error correction.

8) Map author's edit.

9) Error correction.

10) Error-free map stored on a tape file and recorded in the national cartographic file.

11) At this stage a computerized legend can be prepared if it is warranted by the level of interest in the map. The computerized map information can be accessed for interpretation either by means of a computerized legend or a computer generated "Retrieval coding Document" which is manually completed.

After the desired thematic information has been mapped and transferred to the appropriate base map overlay it is checked and edited to conform to national correlation and cartographic standards. The map at this stage consists of an acceptable manuscript for computer input.

Data input is carried out on a digitizing table from a map in manuscript form. Generally, the input phase is done in conjunction with the normal map printing program by manually scribing the thematic lines with an electronic cursor/scriber, and cueing the computer at the start and end points of each line segment. The scribed line segments form an artificial negative for the map printing process at the same time as the x-y coordinates are recorded on disk. Subsequently, the thematic symbols are entered by means of a keyboard. They are positioned within the appropriate area and 'linked' to one of the perimeter line segments. The data are collected and stored on a magnetic disk and transferred to an IBM/370 & 3033 computer located at Datacrown. The IBM processing links the line segments into polygonal areas and associates the areas with their appropriate symbols. This processing phase detects and identifies a number of the errors which may have occurred during the input phase (the error rate is generally low, less than 2%) and produces an error plot and diagnostics. It also generates routine statistics such as symbol frequency, areas, etc. The digitizeo map is updated to eliminate the errors and produce a 'clean' data set of thematic information. A brief description of the organization of data in the Cartographic file is contained in Appendix 11.

The process of transforming thematic lines and symbols on a soil map to digital format requires specialized equipment, software, and personnel. Consequently, only a limited number of data input facilities are planned. In most cases users will submit maps for digitizing to an existing set-up rather than acquiring another cartographic input site and assuming the associated costs.

For this reason and also because the technology and upkeep associated with these procedures are constantly being updated and improved, a complete explanation of cartographic input is not documented in this manual. Full descriptions of the cartographic data input installations including hardware specifications, software, and cartographic/digitizing procedures) are available on request in mimeograph form from CanSIS.

Currently, complete data input systems are operating in Ottawa (4 digitizing tables) and Winnipeg (1 digitizing table). At these installations data are transferred from tables via PDP11 minicomputers to disk and then to tape. The tape is then transferred to an IBM/370 & 3033 for processing and the processed data are plotted on a Calcomp 748 flatbed plotter operating off-line (the Winnipeg installation does not have plotter facilities). The hardware configuration, software and operation at the Ottawa installation is documented in the "CanSIS Cartography Subsystem User's Manual". The installation at Winnipeg is similar but has been modified by their local software personnel.

Data input systems are also operating or under development at Guelph, Saskatoon, Winnipeg and Ottawa in which the data from digitizing tables (1 per installation) is transferred via an IBM Personal Computer to a magnetic disk and then to a streaming tape cassette. These installations

include some proprietary software which allows the map or portions thereof to be displayed on a video screen and copied to hardcopy format. Data input at these installations will be transferred on the tape cassettes for conversion to standard 1600 bpi tape for input to the IBM/370 & 3033.

All data are input by trained cartographers who in most cases have had additional training in input of data to CanSIS by the Cartographic Unit of LRRI. The conventions and procedures for cartographic data input station operators are documented in the "QCU Procedures Manual" (see appendix 1, item 13).

The flow of data into the cartographic file is summarized briefly in Table 5. Once a map has been input checked and declared error free it is accessible for retrieval and interpretation (see section 9.3).

9.2 Cartographic Subsystem Limitations

The cartographic subsystem is designed to be as flexible as possible. However, some of the restrictions of concern are listed below:

9.2.1 Map Configuration: Only rectangular maps are currently acceptable. The system assumes four corners with straight lines between them. This means that maps that do not conform with this must have an "artificial" boundary surrounding them such that all lines and symbols will fit within the rectangle formed by those 4 corners. Insets, Peninsulae that extend across the border, etc. can not be handled, and would necessitate major programming efforts, and perhaps even modifications in the file and record structures.

9.2.2 Size Limitations: There is a limit of 3989 symbols in the map. The number of lines per map is limited to 29419.

A current maximum of 1500 unique symbols is a limitation of the map index (Level 1, see appendix 11). This is becoming a serious limitation for a number of large detailed maps. The effect of increasing capacity for unique symbols is increased costs in all processing steps and in particular in the retrieval step.

Maps up to 36" x 48" can be accommodated. Larger maps are normally split for input but this has no impact on the printed map program. This does mean that the retrieval of computerized maps must be carried out in each portion of the map with a resulting increase in costs.

9.3 Access to Data in the Cartographic File

In simple terms the cartographic subsystem consists of a series of computerized procedures for converting all the information on a map into a series of accurate digital records. In CanSIS this is done by a set of manual procedures which enter into the computer (digitize) the coordinates of each element of line on a map, noting where line segments join, recording the symbol to be associated with each area and associating the symbol with one line segment of the area.

The software associated with the cartographic subsystem performs a variety of checks to verify that every area on the map is closed, that no area contains more than one designation or symbol. As such it provides assistance to the cartographers in their preparation of printed maps. It performs a series of calculations of interest to the authors and users of the maps; such as the area on the map associated with each symbol and the number of areas on the map which have the same symbol. These data are routinely supplied as tabular output to map authors. In association with the printed map program, the Cartographic File produces a listing of the map symbols in machine readable form for direct input into a computerized typesetting unit. It can produce plots of colour separation guides for as many as 15 colours. In addition, the plotter has the capability to produce cartographic quality plots of thematic lines and map symbols. The cartographic file also makes provision for the map to be interpreted (i.e. simplified) for specific applications. From a system standpoint, two basic operations are involved in the interpretation process.

i) The map symbols describing the complete range of characteristics associated with a map area are replaced with a symbol representing or describing only those characteristics pertinent to the interpretation. For any interpretation these data must be specified and input to the computer.

ii) The computer reads the list of symbols and the interpreted symbols which are to replace them and executes data manipulation to prepare an interpreted map. In most cases the interpretation process will result in a reduction in the number of different map symbols. This necessitates a check to make sure that adjacent areas are different - generally when this check is carried out some lines will be removed because the same symbol will apply on both sides of the line. The calculation of area represented by each symbol and the number of occurrences will be repeated for the interpreted symbols. Once these operations are completed the data for the interpretive map are written onto a magnetic tape for plotting.

These two operations represent the processes required within the cartographic subsystem to prepare in interpretive map. In this discussion procedures are described by which the original map symbols are replaced by interpretive ones.

9.3.1 Procedures for Specifying Interpretations

9.3.1.1 Manual Coding Procedures - The Turnaround Document: In the simplest case, the input data for an interpreted map (i.e. the original map symbol and the interpretive symbols which are to replace them) can be prepared manually.

A correspondence list is prepared assigning to each symbol which occurs on the map an appropriate interpretive symbol. This is carried out by generating a computer listing of all symbols occurring on the map and allowing the user to code new interpretive symbols. The listing, "a turnaround document" is then keypunched and the data are input.

The manual procedure can be used to prepare many kinds of interpretations from those which simply select a part of the original symbol such as slope and produce a single-factor-derivative type map to a complex interpretation requiring more or less subjective decisions about the information in the map symbol and other pertinent reference material.

This manual procedure becomes quite tedious where a variety of interpretations for different areas are required, especially on maps or groups of maps containing large numbers of map symbols. A sample of a computer generated "turnaround document" and coding instructions are contained in Appendix 12.

9.3.1.2 Automated Coding Procedures - The Computerized Legend: The computerized legend technique has been developed as an alternative to the manual approach for preparing interpretive maps. It avoids the tedium of preparing a separate interpretation for every map symbol and is somewhat more objective. Before an interpretation can be carried out using a computerized legend, the legend itself must be prepared and defined within CanSIS. Appendix 13.1 contains a summary of the computerized legends currently available.

The computerized legend consists of a computerized array of data which lists each map symbol and all the information necessary to interpret the symbol. In CanSIS computerized legends are prepared as one or a series of RAPID relations; consequently, all the access and data manipulation techniques discussed under the section Data Retrieval from RAPID Relations can be applied to legends. The organization of this table of data is fixed or rigidly structured which means that the location of data (if present) on any characteristics can be specified unequivocably to the computer.

An interpretation is carried out by:

(i) classifying the map symbols into appropriate classes. For this step one or a series of general commands are applied to the structured array of information to effect a grouping of symbols. For example, the series of general commands might be "if the slope class is 0, 1 or 2" and "the parent material is not lacustrine" and "the texture is loam, clay loam, sandy loam or loamy sand" select the map symbol. This step is usually carried out using DREAM to develop the request and then all the selections and groupings are combined in an EASYTRIEVE program.

(ii) Once the map symbols have been grouped into interpretive classes, they must be assigned a new symbol to replace the original ones. This can be done simply by selecting the values of the attribute represented in the interpretation. For example, if a drainage map is being prepared the classification might consist of grouping the map symbols according to their drainage class and the interpretive symbols might simply be the actual drainage class. Otherwise appropriate definitions for the symbols can be prepared so that drainage class 1 would appear on the map as the symbol 'VERY GOOD' and drainage class 2 would be mapped with the symbol 'GOOD' etc. This technique of

defining symbols for the various classes could be used to assign the interpreted symbol 'SUITED'. The EASYTRIEVE program includes this allocation of derived or interpreted symbols to the group of map symbols which had the characteristics of

- slope class in the range 0, 1 or 2
- parent material not lacustrine
- texture of loam, clay loam, sandy loam or loamy sand.

Computerized legends are most effective where large numbers of map symbols are involved and/or where a number of interpretations are required. They are particularly efficient when the map symbols or their components are repetitive, where a legend can be used for several maps, and where interpretations are required on an ongoing basis.

By using a computerized legend, interpretations can be developed in an interactive, iterative fashion (DREAM) by refining the selection and classification request on the basis of tabular data from the legend before presenting it in map form. The major effort both subjective and objective is required at the time the legend is assembled. If the legend is prepared with due consideration for its use, the steps involved in carrying out an interpretation are straightfoward and automated.

Appendix 13.2 lists examples of access to a computerized legend using DREAM and the development of interpretive classes. It further shows how the series of DREAM selections have been combined into an EASYTRIEVE program which also assigns interpretive map symbols and the map legend and title. At this stage most users will contact the CanSIS staff in Ottawa for assistance in preparing the desired interpretive map. Appendix 13.3 contains the form to be submitted for map preparation and illustrate the options available for interpretive map formats.

10 DEVELOPMENT OF NEW CANSIS DATA FILES

10.1 Input Form Development Considerations

To achieve a data base with the lowest possible level of errors, the distance between the data collector and the computer must be kept as short as possible. Ideally, data input by the originator at an intelligent terminal, would give the best results. This assumes the originator has the requird time and skills. Since this rarely happens, computer input forms are developed on which the data may be coded in the field and from which the data may be keyed onto machine readable media.

Computer input forms must be designed for the information to be collected, the circumstances of their use, and the data manipulation required. In the course of the data input procedures, every attempt should be made to make data collection, coding, correction, update and retrieval as flexible as possible.

There are two types of circumstances for which the CanSIS user may want to develop new data input forms.

(1) When an existing data input form accommodates all the information to be collected but is difficult to use because it is not organized in the best fashion for the data being collected or it is too large and comprehensive for specific projects which involve collecting only a small subset of data. The normal approach to deal with these circumstances is to modify and rearrange an existing input form (i.e. "cut and paste") to develop a more acceptable input coding form. Data collected on forms developed in this fashion remain compatible with the complete data file. Consequently existing programs, procedures and computer definitions can be used to manipulate and store the data.

(2) When the data to be collected are not accommodated within one existing data input form, a completely new one is required. Under this circumstance a completely new computer file definition is also required. The development of new forms of this sort should be done in conjunction with the design and analysis of the computer file and its layout. It should proceed cooperation and consultation with the CanSIS staff.

10.1.1 Modifying Existing Data Forms

Under conditions where an acceptable form can be developed by rearranging an existing form and using only the parts required the following considerations apply:

(1) The key fields (the data fields normally located at the beginning of the 80 column record which serve to uniquely identify each line of data) must be retained and filled in according to the rules of the original form e.g. numeric fields can have only valid numbers.

(2) The card types and card numbers must be retained; the input data fields (variables) not required on each cardd may be deleted. Card integrity must be maintained i.e. data input on different cards in the original data form cannot be combined onto a single card.

(3) The column position of fixed format or positional entries must be retained.

10.1.2 Developing Completely New Data Forms

Computer input forms for data not accommodated within existing CanSIS files should be developed after the types of analysis and output have been specified. This will ensure that the data are collected and input in a format that facilitates the normal manipulations required. In addition, based on experience gained in developing CanSIS files and associated input forms the following points should be considered:

1) Initial destination of the input data may be either the custom developed CanSis file management system (FMS) or the RAPID data base management system (DBMS). For data which will be collected over a long time and for which good error correction and update procedures are required the FMS is recommended. For short term projects or projects in which data is not added on an ongoing basis the data may be directly input to the RAPID DBMS. There is no rigid guideline regarding where the data are input; however for data which are collected and input in batches the FMS facilitates the organization and archive processes.

2) The styles of data entry depend on the type of property being characterized, the way the property is to be used and the frequency with which it is collected and used. If the data are input to the FMS, the four styles of input are fixed entry (either character or decimal), self-defining entry, value-coded entry, and composite entry as described in section 5..2.1. When data are input to the RAPID DBMS, only the fixed entry style of data entry is possible. Because all data input to the FMS is normally reformatted and defined to RAPID for manipulation and output, it is recommended that regardless of whether the data are input to the FMS or RAPID that only fixed entries be used. In this way the amount of reformatting required to define the data from the FMS to the RAPID DBMS is minimized.

3) The definition and organization of the key fields in the data (the character sequence which uniquely defines each reecord or line to the computer system) is a very important part of file and input form design. It is essential for data input, editing and error correction. The key is normally located in the first columns of each 80 column record. If more than 80 columns are required for a complete record then the key will normally contain a counter in the form of a card type and/or number designation. The types of information (fields or variables)

included in the key should be as similar as possible to keys of other CanSIS file. Normally lthey will include the province, year, project or site identification number, and initials of the person recording the information. Commonly the key will also include the date, horizon or depth, or sample type. Frequently, more than one length of key will be required for the same input record e.g. records which contain both site characteristics and horizon characteristics. In these cases the data will be split into different files according to key and the common elements of the key will allow the data to be recombined on output.

4) In designing the data fields on a new computer input form, it is most convenient to have a separate field for each characteristic. In practice it is almost impossible to anticipate the range of characteristics which the user may wish to record. In addition, there are normally some types of characteristics which occur only occasionally and will most often be left blank. If the attempt is made to produce a computer form which encompasses all possible characteristics the result is a very large and extensive form in which many fields will contain no data. Furthermore it will quickly become inadequate as users wish to record data for which no input field has been provided. Adding fields requires the costly steps of changing the input forms and the computer definitions and procedures.

   Experience with other CanSIS input forms suggests that the most desirable organization for the data fields is as follows:

   (a) specific fields to be defined for characteristics which are routinely collected.
   (b) paired fields are defined to allow for definition of the property being measured and the value observed.

   Examples of this type of field may be found on pages 21, or 27 of the performance management form. These fields are used for characteristics which are only accosionally measured or to accommodate the measurement of properties not anticipated when the form was developed. This approach means that a very flexible program of data collection is possible without changes in the input form or the computer definition.

5) Integration or linking fields should be incorporated into the computerized record whenever feasible. These are fields which have standard definitions and values in CAnSIS and allow data in one file to be linked to another file avoiding the unnecessary duplication of information. For example the inclusion of the standard 3 character soil code in a computerized legend facilitates a merge between the legend and the Soil Names File.

6) In a developing data file it may on occasion be desirable to define some standard fields not allocated to any specific property. They may be used later to indicate special

properties, changes to the site, or project specific data. If this capability is desired it can be conveniently implemented in association with a special notes line.

7) Organization of the form is very important. In general terms, the sequence of data elements should follow the logical progression of examination of the site, soil or item. This will reduce the amount of flipping back and forth through the form.

8) It has been found very useful to include a free format section on any form that is designed. This is useful not only for comments, but can also be used for extra data items previously overlooked, added on, etc.

9) It is essential that the data collection form be developed at the same time as the instruction manual to use the form. If the two progress at the same time, the needs of the potential user will become evident. In addition, both documents should be circulated to all intended and potential users before use. Even when active use begins, it should be provisional for at least a field season. Although forms always seem to be in a state of perpetual change, the ramifications of change affect all parts of the data collection system.

10) A well designed form is a compromise between the needs and requirements of the data collector in the field and the needs and requirements of the machine and the well being of the keypunch operator. For this reason, the following points should be kept in mind.

    1. The form should be easy to use.
    2. It should allow for collection of all pertinent data.
    3. It should be easy to keypunch.

10.1.3 Keying Instructions

Data documents that have been filled in during the field survey are generally converted into machine readable code by commercial keypunch operators who know nothing about the data. To ensure accurate interpretation of the document by the operator, detailed instructions must accompany the data documents.

Although the specific detail of the coding documents will vary with each document type, the following example constitutes the keypunch instruction for the Performance/Management file.

This form consists of up to 30 pages (14 x 8.5) of which only 6 are of similar format. Of these, 4 pages are entirely free-format entries; 20 are entirely fixed column positional entries; and the rest are a combination of these on the same page.

10.1.3.1 Fixed-Column Entries:

- Each page is self-defining. That is, all data to be keyed from it on that page.

- All punched cards must start with the 13 to 23 columns of data normally found in the top left-hand corner of each page.
- The length of these "prefix" data and the number of cards vary from page to page (and from form to form).
- Key only the lines for which data are entered. The order in which the cards are keyed is unimportant.

10.1.3.2 Free-Format Entries:

- Key only circles and box-filled entries on 6 or 12 column tab boundaries. Entries are of three types.

On form	Keyed as	
C201**	'C201 '	6 columns
C601**+-1 ++	'C601 +-1 ++'	12 columns
07 03*	'07 03 '	6 columns

- Columns 1 to 15 are duplicated for each card required.
- Starting in column 16 to 75, in any order, key as many entries as possible per card. Do not, however, split entries between cards.
- The *'s and unfilled boxes are keys as blanks.

10.2 File Definition Considerations

As the data colllection form is developed, the structure, location and names of all the fields have to be communicated to the computer through the file definition program. The purpose of this program is to define the information on the input document in a format suitable for the computer. EAch piece of information is identified and described in a logical manner. In order to define a data file, the user must describe the format of the data as it is stored in a record and as it is entered via machine-readable form.

For data being defined to the CanSIS file management system, details of the file definition procedure are contained in the File Management Subsystem (FMS) User's Manual.

For data being defined to the RAPID data base management system, details of the definition procedure are contained in the RAPID documentation and an example of its application in CanSIS is available in the internal report Systems Operation at Datacrown using an Account Independent of Agriculture Canada.

The file definition process is a valuable tool in verifying a data collection form. The computer does a number of validity checks helping to make the form logically correct.

10.3 Computer Checking:

Editing of data is one of the most time consuming and laborious aspects of data systems. However, data quality is largely a function of effective editing. In CanSIS, the computer does a number of tests which check data conformity to the data definition. A field defined to be numeric, must contain a numeric value. The value of numeric fields must be within the specific range. Both parts of a composite entry must be present. A self-defining entry name must be valid (i.e. it must fall in the designated range for each field). With some modification, the system could be modified to check that all alphanumeric values are left justified and that all numeric values should be zero filled. In addition, with relatively little work the system could also check the data to ensure that codes are valid and that for composite fields, one portion missing, would cause a warning message to be printed.

The generation of consistency lchecks aree the responsibility of the user who uses them for checking against guidelines, standards or classifications. These may involve the examination of one, two or more fields for data accuracy and logical consistency. This could take the form of: if A = B and C = D and E = F or E = G then the soil is a Podzol; e.g. if Ae is less than 15% clay, and the B horizon is greater than % clay of AE + 3% then horizon is a Bt. Other desirable checks would include testing if a field contains data. The system may also test for a compatible number of occurrences of a repeating segment. For example, raise an error condition if the number of horizons specified in the analytical section exceeds the number of horizons specified in the descriptive section or, check that the number of values specified does not exceed the defined limit.

Consisting checks of the type referred to in the preceeding paragraph are file and field specific and depend on the semantics of the data. Consequently, the users must specify what tests are requ;ired and specific programs or special editing capabilities are then developed.

APPENDIX 1

CanSIS Documentation (1983)

1. The Canada Soil Information System (CanSIS) Manual for describing soils in the field (1982 Revised). J.H. Day editor. L.R.R.I. contribution number 82-52. Research Branch, Agriculture CAnada, Ottawa, Ontario 1983.

2. Output Manual for Soil Data File (Detail) (in preparation).

3. Canada Soil Information System (CanSIS) File Management Subsystem (FMS) User's Manual Version 1.0, October 1978. B. Gibb & A. Desrochers, Systems & Consulting Directorate, Agriculture Canada, Ottawa, Ontario. (Manuscript - publication not planned).

4. The Canada Soil Information System (CanSIS) Manual for Describing Performance/Management Data. K.B. MacDonald & Z.S. Strzelczyk. L.R.R.I. contribution number 108. Research Branch, AGriculture Canada, Ottawa, Ontario 1981.

5. User's Output Manual for the Soil Performance & Management File of CanSIS. K.B. MacDonald & Z.S. Strzelczyk. Research Branch, Agriculture Canada, Ottawa, Ontario. (In draft form).

6. The Land Potential Data Base for Canada: User's Handbook. V. Kirkwood, J. Dumanski, K.B. MacDonald, R.B. Stewart, and B. Gronas. L.R.R.I. contribution number 82-68, Research Branch, Agriculture Canada, Ottawa, Ontario 1983.

7. CanSIS Cartography Subsystem User's Manual (version 1.0) November 1978. J. Buchanan and D. Regan. Systems & Consulting Directorate, AGriculture CAnada, Ottawa, Ontario. (Manuscript - publication not planned).

8. Spatial Display Point Plot Package: User Manaual and Examples. V. Brennan, J. Buchanan, K.B. MacDonald, and Z.S. Strzelczyk. Research Branch, Agriculture Canada, Ottawa, Ontario. (Manuscript).

9. Map Interpretations and Computerized Legends: A preliminary manual for the Natural Resource Sciences. B. Kloosterman and K.B. MacDonald. Research Branch, Agriculture Canada, Ottawa, Ontario. (Manuscript).

10 Wetlands Registry Manual (in preparation). Research Branch, Agriculture Canada, Ottawa, Ontario.

11. Systems Operation at Datacrown using an Account Independent of Agriculture Canada: Some Considerations for CanSIS Clients Using Procedures Develped & Maintained by Agriculture Canada and Using DAta Compiled & Stored by Agriculture Canada. Internal L.R.R.I. Report, September 1982. K.B. MacDonald & A.J. Sunnak.

12. The Canada Soil Information System CanSis File Retrievals (Samples) B. Lacelle. Research Branch, Agriculture Canada, Ottawa, Ontario. (Manuscript).

13. QCU Procedures Manual. Cartographic Unit, Research Branch, Agriculture Canada, Ottawa, Ontario.

Commercial Documentation of Software and Procedures used in CanSIS.

1. Datacrown Inc. Systemguide Toronto Systemcentre ZCSG - 1982.

2. EASYTRIEVE Reference Manual, Pansophic Systems Inc., 709 Enterprise Drive. Oak Brook, Illinois.

3. The SAS Users Guide, ed. K.A. Council & J.T. Helwig, SAS Institute Inc. Raleigh, North Carolina.

4. RAPID Data Base Management General Information (contains references to the entire range of RAPID documentation). Special Resources Subdivision, Systems Development Division, Statistics Canada, 12-P R.N. Coats Building, Tunney's Pasture, Ottawa, Ontario, K1A 0C6.

APPENDIX 2: ACCESSING CANSIS (October 1983)

2.1 Sample Logon

The following is a simulated dialogue between the User and DATAPAC. Following conventional pracice the computer prompting will be given in upper case while the user response is in lower case.

To activate DATAPAC, dial xxx-xxxx (phone numbe of the closes DATAPAC port) and, when the computer tone is heard, place the receiver in the modem. The system should be set to full duplex with a baud rate of 300. It will be activated by typing.

```
 3.  USER:       ...(CR)
 4.  TERMINAL:   DATAPAC:  TERMINAL ADDRESS
 5.  USER:       40200027(CR)
                     OR '40200058(CR)')
 6.  TERMINAL:   DATAPAC: CALL CONNECTED
 7.  USER:       (CR)
 8.  TERMINAL:   TOR INVALID SITE OR APPLICATION PLEASE REENTER
 9.  USER:       sna
10.  TERMINAL:   ENTER FUNCTION:
11.  USER:       tsol userid/password
12.  TERMINAL:   LOGON TO TORTSOl IN PROGRESS
                 USERID LOGON IN PROGRESS AT HH:MM:SS ON MNTH, YEAR
                 NO BROADCAST MASSAGES
                 READY
```

You are now singed on to the DATACROWN Toronto System 1 TSO, for more information on DATAPAC 3101 and the DATAPAC number in your area, the following document is available through the DATACROWN online guide system by submitting:

```
//guide exec guide,subject=dpac3101
```

NOTE: For those who use TSO through the agrinet communications system or who do not use DATAPAC access, only steps 8 to 12 in the above procedure apply. The procedure preceding these steps does not change.

Example of Signing onto Datacrown Through Datapac

```
...(CR)
DATAPAC: 2040 0070
4020027

DATAPAC:  CALL CONNECTED

TOR INVALID SITE OR APPLICATION PLEASE REENTER
sna

NATIVE MODE
ENTER FUNCTION:  tsol ag230??/??

DSC1061 LOGON TO TORTSO2 IN PROGRESS

ENTER ACCOUNT NUMBER -
ag230-1200

ENTER PROGRAMMER NAME
name

AG230BK LOGON IN PROGRESS AT 098:02:17 ON OCTOBER 24, 1983
20.35.03 JOB 6422 $HASP165 ID376 ENDED AT TORONTO
READY
LOGOFF

AG230BK LOGGED OFF TSO AT 09:02:31 ON OCTOBER 24, 1983.
```

2.2 Accessing RAPID Relations

Once the computer logon is complete the user is accessing the system in TSO (time sharing option). This is indicated by the key work:
READY

The software package DREAM may be accessed by typing one of the following CLIST procedures:

ex sis01(lsisdrm) - for RAPID RELATIONS for the Land Potential Data Base and Computerized Legends

ex sis01(lsissoil) - for RAPID RELATIONS developed for Soil Survey

ex sis01(lsisappl) - for RAPID RELATIONS developed for Wetlands Registry and Performance and Management

ex sis01(lsisgend) - general access

The information listed for NEW USERS for each of these commands is listed on the following pages.

CLIST:LSISGEND

WELCOME, YOU HAVE JUST INVOKED THE PACKAGE DREAM, WHICH ALLOWS ON-LINE ACCESS TO RAPID RELATIONS WHICH FORM PART OF CANSIS.
IF YOU HAVE ANY PROBLEMS, QUESTIONS OR SUGGESTIONS CALL:

PETER BRIMACOMBE (613)593-7791

2.2.2 Land Potential Data Base and Computerized Legends

CLIST: LSISDRM

```
WELCOME TO THE WONDERFUL WORLD OF CANSIS JUNE/84.
IF YOU HAVE ANY PROBLEMS, QUESTIONS OR SUGGESTIONS CALL:
      PETER BRIMACOMBE AT (613) 593-7791 OTTAWA

RAPID RELATIONS FOR LAND POTENTIAL DATA BASE AND COMPUTERIZED LEGENDS.

     SOIL        LAND POTENTIAL DATA BASE
     E100        GENERALIZED SOIL LANDSCAPE (PRAIRIES)

     E026        PEI LEGEND  - SYMBOLS
     E027                    - SOIL SERIES
     E028                    - DECODED SYMBOLS
     E029                    - FULLY EXPANDED

     E050        ONT LEGEND  - HURON, YORK, ONTARIO COUNTIES
     E051                    - WATERLOO - SOIL SYMBOL
     E052                    - WATERLOO - MAP SYMBOL

     E080        N.B. LEGEND

     E060        SASK LEGEND - MAP SYMBOL
     E061                    - MAP UNIT CHARACTERISTICS
     E062                    - MAP UNIT COMPOSITION
     E063                    - SERIES

     NOTES:      (1)  ARCHIVE
                      ANY RELATION OR DICTIONARY WHICH IS NOT ACCESSED FOR
                      MORE THAN 60 DAYS MAY BE ARCHIVED.  IF THE RELATION IS
                      UNAVAILABLE, THIS CLIST WILL ABEND TO THE 'READY' MODE.
                      TO CONFIRM WHETHER OR NOT THE RELATION IS AVAILABLE,
                      FROM THE 'READY' STATE, TYPE
                           LDS SIS01.XXXX
                           WHERE XXXX IS THE REELATION NAME
                      IF THE RELATION IS ARCHIVED, CONTACT PETER BRIMACOMBE AT
                      (613) 593-7791, OTTAWA TO HAVE IT RESTORED.

                 (2)  DATA DICTIONARIES
                      SIS01.FAOD.DICT  CONTAINS SOIL
                      SIS01.EXPR.DICT  CONTAINS E100
                      SIS01.EX10.DICT  CONTAINS E026, E027, E028, E029
                      SIS01.EX05.DICT  CONTAINS E050
                      SIS01.EX55.DICT  CONTAINS E051, E052
                      SIS01.EX08.DICT  CONTAINS E080
                      SIS01.EX03.DICT  CONTAINS E060, E061, E062, E063

     RELATIONS AVAILABLE JUNE/84:  SOIL, E100, E051, E052, E060
```

2.2.3 Soil Survey

CLIST: LSISSOIL

```
WELCOME TO THE WONDERFUL WORLD OF CANSIS JUNE/84.
IF YOU HAVE NAY PROBLEMS, QUESTIONS OR SUGGESTIONS CALL:
     PETER BRIMACOMBE AT (613) 593-7791 OTTAWA

RAPID RELATIONS DEVELOPED FOR SOIL SURVEY

     SL1E    SOIL NAMES  -  ENGLISH
     SL1F                -  FRENCH

     SD01    SOIL DATA (DETAIL)  -  SITE
     SD02                        -  HORIZON
     SD03                        -  NOTE

     MT02    MONITOR RELATION  -  MONITORING OR TIME SERIES DATA

     MN01    MANITOBA DAILY  -  SITE
     MN02                    -  HORIZON
     MN03                    -  NOTES

     NOTES: (1)  ARCHIVE
                 ANY RELATION OR DICTIONARY WHICH IS NOT ACCESSED FOR MORE
                 THAN 60 DAYS MAY BE ARCHIVED!  IF THE RELATION IS
                 UNAVAILABLE, THIS CLIST WILL ABEND TO THE 'READY' MODE.  TC
                 CONFIRM WHETHER OR NOT THE RELATION IS AVAILABLE, FROM THE
                 'READY' STATE, TYPE
                         LDS SIS01.XXXX
                         WHERE XXXX IS THE RELATION NAME
                 IF THE RELATION IS ARCHIVED, CONTACT PETER BRIMACOMBE AT
                 (613) 593-7791, OTTAWA TO HAVE IT RESTORED.

                (2) DATA DICTIONARIES
                 SIS01.SLDE.DICT  CONTAINS SL1E
                 SIS01.SLDF.DICT  CONTAINS SL1F
                 SIS01.SLDD.DICT  CONTAINS SD01, SD02, SD03, MT02
                 SIS01.MNDD.DICT  CONTAINS MN01, MN02, MN03

     RELATIONS AVAILABLE JUNE/84:  SL1E, SL1F, SD01, SD02, MT02, MN01,
                                   MN02, MN03
```

2.2.4 Wetlands Registry and Performance Management

CLIST: LSISAPPL

WELCOME TO THE WONDERFUL WORLD OF CANSIS JUNE/84.
IF YOU HAVE ANY PROBLEMS, QUESTIONS OR SUGGESTIONS CALL:
PETER BRIMACOMBE AT (613) 593-7791 OTTAWA

RAPID REELATIONS DEVELOPED FOR THE SOIL PERFORMANCE AND MANAGMENT FILE AND THE WETLANDS REGISTRY

PMA1	PM	- SITE
PMB1		- GLOBAL MANAGEMENT
PMC1		- WEATHER
PMD1		- HORIZON
PME1		- FACTOR
PMF1		- FACTOR - LEVEL
PMG1		- TREATMENT
PMH1		- SOIL ANALYSIS
PMI1		- CROP YIELD AND QUALITY
PMI2		- CROP DEVELOPMENT
PMI3		- NOTES

WTL1	WETLANDS	- AREA
WTL2		- SITE
WTL3		- HYDROLOGY, MORPHOLOGY, PHYSICAL AND CHEMICAL ANALYSIS
WTL4		- VEGETATION
WTL5		- NOTES

NOTES: 1) ARCHIVE
ANY RELATION OR DICTIONARY WHICH IS NOT ACCESSED FOR MORE THAN 60 DAYS MAY BE ARCHIVED. IF THE RELATION IS UNAVAILABLE, THIS CLIST WILL ABEND TO THE 'READY' MODE. TO CONFIRM WHETHER OR NOT THE RELATION IS AVAILABLE, FROM THE 'READY' STATE, TYPE
LDS SIS01.XXXX
WHERE XXXX IS THE RELATION NAME
IF THE RELATION IS ARCHIVED, CONTACT PETER BRIMACOMBE AT (613) 593-7791, OTTAWA TO HAVE IT RESTORED!

2) DATA DICTIONARIES
SIS01.PMDD.DICT CONTAINS THE P/M RELATIONS
SIS01.WTLD.DICT CONTAINS THE WETLANDS RELATIONS

RELATIONS AVAILABLE JUNE/84: PMA1, PMB1, PMC1, PMD1, PME1, PMG1, PMH1, PMI1, PMI2, PMI3, WTL1, WTL2, WTL3, WTL4, WTL5

CanSIS SOIL NAMES FILE

PROV. | SOIL CODE
1 | 3

CARD NO. 0001

SOIL NAME

01 ____ **

REPORT NO.

02 ____ *

PROXIMITY TO U.S.A.

C801** Within specified limit
C802** Not within spec. limit

STATUS

Acceptance

C811** Reserved
C812** Open for mapping
C813** Closed for mapping

LEVEL

C821** Series
C822** Single M.U.
C823** Compound M.U.1
C824** Compound M.U.2
C825** Compound M.U.3
C826** Compound M.U.4

SOIL CLASSIFICATION OF DOMINANT PEDON

Brunisol

E201** O.MB
E202** E.MB
E203** GL.MB
E204** GLE.MB

E211** O.EB
E212** E.EB
E213** GL.EB
E214** GLE.EB

E221** O.SB
E222** E.SB
E223** DU.SB
E226** FR.SB
E224** GL.SB
E225** GLE.SB
E231** O.DYB
E232** E.DYB
E233** DU.DYB
E236** FR.DYB
E234** GL.DYB
E235** GLE.DYB

Chernozemic

E241** O.B
E242** R.B
E243** CA.B
E244** E.B
E245** SZ.B

E246** GL.B
E247** GLR.B
E248** GLCA.B
E249** GLE.B
E250** GLSZ.B

E261** O.DB
E262** R.DB
E263** CA.DB
E264** E.DB
E265** SZ.DB
E266** GL.DB
E267** GLR.DB
E268** GLCA.DB
E269** GLE.DB
E270** GLSZ.DB

E281** O.BL
E282** R.BL
E283** CA.BL
E284** E.BL
E285** SZ.BL
E286** GL.BL
E287** GLR.BL
E288** GLCA.BL
E289** GLE.BL
E290** GLSZ.BL
E291** O.DG
E292** R.DG
E293** CA.DG
E294** SZ.DG
E295** GL.DG
E296** GLR.DG
E297** GLCA.DG
E298** GLSZ.DG

Cryosolic

E301** O.TC.
E302** BR.TC
E303** R.TC
E304** GL.TC
E311** O.SC
E312** BR.SC
E313** R.SC
E314** GL.SC

E321** FI.OC
E322** ME.OC
E323** HU.OC
E324** TFI.OC
E325** TME.OC
E326** THU.OC
E327** GC.OC

Gleysolic

E331** O.HG
E332** R.HG
E333** FE.HG

E341** O.G
E342** R.G
E343** FE.G

E351** O.LG
E352** HU.LG
E353** FE.LG
E354** FR.LG

Luvisolic

E361** O.GBL
E362** BR.GBL
E363** PZ.GBL
E364** GL.GBL
E365** GLBR.GBL
E366** GLPZ.GBL

E371** O.GL
E372** D.GL
E373** BR.GL
E374** PZ.GL
E375** SZ.GL
E376** FR.GL
E377** GL.GL
E378** GLD.GL
E379** GLBR.GL
E380** GLPZ.GL
E381** GLSZ.GL
E382** GLFR.GL

Organic

E401** TY.F
E402** ME.F
E403** HU.F
E404** LM.F
E405** CU.F
E406** T.F
E407** TME.F
E408** THU.F
E409** HY.F

E411** TY.M
E412** FI.M
E413** HU.M
E414** LM.M
E415** CU.M
E416** T.M.
E417** TFI.M
E418** THU.M
E419** HY.M

E421** TY.H
E422** FI.H
E423** ME.H
E424** LM.H
E425** CU.H
E426** T.H
E427** TFI.H
E428** TME.H
E429** HY.H

E431** TY.FO

Podzolic

E441** O.HP
E442** OT.HP
E443** P.HP
E444** DU.HP
E445** FR.HP

E451** O.FHP
E452** OT.FHP
E453** P.FHP
E454** DU.FHP
E455** FR.FHP
E456** LU.FHP
E457** SM.FHP
E458** GL.FHP
E459** GLSM.FHP
E460** GLOT.FHP
E461** O.HFP
E462** OT.HFP
E463** P.HFP
E464** DU.HFP
E465** FR.HFP
E466** LU.HFP
E467** GL.HFP
E468** GLSM.HFP
E469** SM.HFP
E470** GLOT.HFP

Regosolic

E501** O.R
E502** CU.R
E503** GL.R
E504** GLCU.R

E511** O.HR
E512** CU.HR
E513** GL.HR
E514** GLCU.HR

Solonetzic

E521** B.SZ
E522** DB.SZ
E523** BL.SZ
E524** A.SZ
E525** GLB.SZ
E526** GLDB.SZ
E527** GLBL.SZ

E531** B.SS
E532** DB.SS
E533** BL.SS
E534** DG.SS
E535** G.SS
E536** GLB.SS
E537** GLDB.SS
E538** GLBL.SS
E539** GLDG.SS
E540** GLG.SS

E551** B.SO
E552** DB.SO
E553** BL.SO
E554** DG.SO

E555** G.SO
E556** GLB.SO
E557** GLDB.SO
E558** GLBL.SO
E559** GLDG.SO
E560** GLG.SO

Phases

E571** Carbonated
E572** Lithic
E573** Peaty
E574** Saline

PARENT MATERIAL

I	II	III	
			Physical Component
A101**	A601**	B101**	Undifferentiated
A102**	A602**	B102**	Fragmental
A114**	A614**	B114**	Sandy skeletal
A115**	A615**	B115**	Loamy skeletal
A116**	A616**	B116**	Clayey skeletal
A104**	A604**	B104**	Coarse loamy and coarse silty
A105**	A605**	B105**	Fine loamy and fine silty
A106**	A606**	B106**	Clayey
A107**	A607**	B107**	Stratified (mineral)
A108**	A608**	B108**	Stratified (min. and organic)
A109**	A609**	B109**	Sandy
A110**	A610**	B110**	Loamy
A111**	A611**	B111**	Fibric
A112**	A612**	B112**	Mesic
A113**	A613**	B113**	Humic
			Chemical Component
A201**	A701**	B201**	Undifferentiated
A202**	A702**	B202**	Extremely/strongly acidic
A203**	A703**	B203**	Medium acid/neutral
A204**	A704**	B204**	Weakly calcareous
A205**	A705**	B205**	Moderately/very strongly calc.
A206**	A706**	B206**	Extremely calcareous
A207**	A707**	B207**	Calcareous and saline
			Mode of deposition or accumulation
A301**	A801**	B301**	Fluvial
A302**	A802**	B302**	Colluvial
A303**	A803**	B303**	Eolian
A304**	A804**	B304**	Fluvioeolian
A305**	A805**	B305**	Fluviolacustrine
A306**	A806**	B306**	Fluviomarine
A307**	A807**	B307**	Glaciofluvial
A308**	A808**	B308**	Glaciolacustrine
A309**	A809**	B309**	Glaciomarine
A310**	A810**	B310**	Lacustrine
A311**	A811**	B311**	Lacustro-till
A312**	A812**	B312**	Marine
A313**	A813**	B313**	Morainal (till)
A314**	A814**	B314**	Organic material (undiff.)
A315**	A815**	B315**	Sphagnum
A316**	A816**	B316**	Brown moss-fen
A317**	A817**	B317**	Sedge-fen
A318**	A818**	B318**	Woody-fen
A319**	A819**	B319**	Woody-forest
A320**	A820**	B320**	Feather moss-forest
A321**	A821**	B321**	Lichen-forest
A322**	A822**	B322**	Aquatic
A323**	A823**	B323**	Mixed peat
A324**	A824**	B324**	Coprogenous earth
A325**	A825**	B325**	Marl
A326**	A826**	B326**	Diatomaceous earth
A327**	A827**	B327**	Residual
A328**	A828**	B328**	Bedrock
A329**	A829**	B329**	Intrusive
A330**	A830**	B330**	Extrusive
A331**	A831**	B331**	Sedimentary
A332**	A832**	B332**	Metamorphic
A333**	A833**	B333**	Ice

LANDFORM CLASSIFICATION

Surface Expression

B701** Apron
B702** Blanket
B703** Fan
B704** Hummocky
B705** Inclined
B706** Level
B707** Rolling
B708** Ridged
B709** Steep
B710** Terraced
B711** Undulating
B712** Veneer
B713** Bowled
B714** Domed
B715** Floating
B716** Horizontal
B717** Plateau
B718** Ribbed
B719** Sloping
B720** Pitted
B721** Patterned

MINERAL SOIL FAMILY CRITERIA FOR THE DOMINANT PEDON

Particle Size of control section

E601** Fragmental
E602** Fragmental/loamy skeletal
E603** Fragmental/clayey skeletal
E604** Fragmental/sandy
E605** Fragmental/loamy
E606** Fragmental/clayey

E611** Sandy skeletal
E612** Sandy skeletal/clayey skeletal
E613** Sandy skeletal/loamy
E614** Sandy skeletal/clayey

E621** Loamy skeletal
E622** Loamy skeletal/fragmental
E623** Loamy skeletal/clayey

E631** Clayey skeletal
E632** Clayey skeletal/fragmental
E633** Clayey skeletal/sandy skeletal
E634** Clayey skeletal/sandy

E641** Sandy
E642** Sandy/clayey skeletal
E643** Sandy/loamy
E644** Sandy/clayey

E651** Loamy
E652** Coarse loamy
E653** Fine loamy
E654** Coarse silty
E655** Fine silty

E661** Loamy/fragmental
E662** Loamy/sandy skeletal
E663** Loamy/sandy
E671** Coarse loamy/fragmental
E672** Coarse loamy/sandy skeletal
E673** Coarse loamy/clayey
E681** Coarse silty/fragmental
E682** Coarse silty/sandy skeletal
E683** Coarse silty/sandy
E684** Coarse silty/clayey

E691** Fine loamy/fragmental
E692** Fine loamy/sandy skeletal
E693** Fine loamy/sandy
E694** Fine loamy/clayey

E701** Fine silty/fragmental
E702** Fine silty/sandy skeletal
E703** Fine silty/sandy
E704** Fine silty/clayey

E711** Clayey
E712** Fine clayey
E713** Very fine clayey

E721** Clayey/fragmental
E722** Clayey/sandy skeletal
E723** Clayey/loamy skeletal
E724** Clayey/sandy
E725** Clayey/loamy

E731** Fine clayey/fragmental
E732** Fine clayey/sandy skeletal
E733** Fine clayey/loamy skeletal
E734** Fine clayey/sandy
E735** Fine clayey/loamy

E741** Very fine clayey/fragmental
E742** Very fine clayey/sandy skeletal
E743** Very fine clayey/loamy skeletal
E744** Very fine clayey/sandy
E745** Very fine clayey/loamy

E751** Cindery
E752** Cindery/sandy skeletal
E753** Cindery/sandy
E754** Cindery/loamy

E761** Ashy
E762** Ashy/cindery
E763** Ashy/loamy skeletal
E764** Ashy/loamy

E765** Ashy skeletal

E771** Thixotropic
E772** Thixotropic/fragmental
E773** Thixotropic/sandy skeletal
E774** Thixotropic/loamy skeletal
E775** Thixotropic/sandy
E776** Thixotropic/loamy

E781** Thixotropic skeletal
E791** Grumic

Mineralogy of control section

E801** Carbonatic
E802** Serpentinitic
E803** Gypsic
E804** Micaceous
E805** Siliceous
E806** Mixed nonclay
E807** Kaolinitic
E808** Montmorillonitic
E809** Illitic
E810** Vermiculitic
E811** Chloritic
E812** Mixed clay
E813** Sulfureous

Soil Depth

E831** Extremely shallow lithic
E832** Very shallow lithic
E833** Shallow lithic
E834** Extremely shallow cryic
E835** Very shallow cryic

Reaction of C

Do not use for Chernozemic, Solonetzic, .GBL, .MB or .EB

E841** Acid
E842** Neutral
E843** Alkaline

Calcareousness of C

E851** Weakly calcareous
E852** Strongly calcareous
E853** Extremely calcareous

Soil Temperature

E861** Extremely cold
E862** Very cold
E863** Cold
E864** Cool
E865** Mild

Soil Moisture

E871** Peraquic
E872** Aquic
E873** Subaquic
E874** Perhumid
E875** Humid
E876** Subhumid
E877** Semiarid
E878** Subarid
E879** Arid

ORGANIC SOIL FAMILY CRITERIA FOR THE DOMINANT PEDON

Organic Surface tier

F001** Fennic
F002** Silvic
F003** Sphagnic
F004** Mesic
F005** Humic

Mineral Surface tier 15-40 cm thick

F011** Sandy
F012** Coarse loamy
F013** Coarse silty
F014** Fine loamy
F015** Fine silty
F016** Clayey

Reaction of control section

F021** Euic
F022** Dysic

Soil Temperature

E861** Extremely cold
E862** Very cold
E863** Cold
E864** Cool
E865** Mild

Soil Moisture

E871** Peraquic
E872** Aquic
E873** Subaquic
E874** Perhumid
E875** Humid

Particle Size of Terric Layer

F051** Fragmental
F052** Sandy
F053** Sandy skeletal
F054** Loamy
F055** Loamy skeletal
F056** Clayey
F057** Clayey skeletal

Limnic Materials

F061** Mari
F062** Diatomaceous
F063** Coprogenous

Depth to lithic contact

F071** Shallow < 100 cm
F072** Mod. deep 100-160 cm

03 SUBCLIMATE

APPENDIX 5: Example DREAM Output from the Soil Names File

Example 1

```
?:
prov = 05
?;
parszcs = s
?;
surfexp = level,hummocky
?;
display prov soilcode soilname classdp pmlphyco pmlcheco pmldep
?;
go
RETRIEVAL TO USE SCANNING.
(RETRIEVAL BEGINS;)
```

PROV	SOILCODE	SOILNAME	CLASSDP	PM1PHYCO	PM1CHECO	PM1MDEP
	1000 ROWS, NO HITS					
	2000 ROWS, NO HITS					
05	GNY	GRANBY	O.HG	SANDY	MOD/VSTR-CALC	FLUVIAL
05	GVR	GOUVEREAU	O.HG	SANDY	MEDIUM-ACID/NEUT	FLUVIAL
05	KEK	KENABEEK	O.G	NO-INFO	NO-INFO	NO-INFO
05	WDG	WENDIGO	O.HFP	SANDY	NO-INFO	GLACIOFLUVIAL
	1000 ROWS, NO HITS					

```
(RETRIEVAL COMPLETE)
ENTER 'END' OR NEXT REQUEST
```

Example 2

```
?:
prov = 01
?:
count soilname by surfexp
?:
go
RETRIEVAL TO USE SCANNING.
(RETRIEVAL BEGINS:)
COUNT SOILNAME            999          1000 HITS.
         1000 ROWS, NO HITS
         2000 ROWS, NO HITS
FINAL COUNT SOILNAME BY SURFEXP
------------------------------
APRON                       6
BLANKET                   348
DOMED                       1
FAN                        41
FLOATING                    1
HORIZONTAL                  3
HUMMOCKY                   25
INCLINED                   21
LEVEL                     134
NO-INFO                   947
RIDGED                     10
ROLLING                    32
STEEP                      13
TERRACED                   95
UNDULATING                 28
VENEER                    171
* FINAL COUNT *          1876
'END' OR RETURN

(RETRIEVAL COMPLETE)
ENTER 'END' OR NEXT REQUEST
?:
```

Example 3

```
?:
prov = 06
?:
soilcode > szz
?:
display prov shrtname classdp mincs parszcs soiltemp soilmois
?:
go
RETRIEVAL TO USE SCANNING.
(RETRIEVAL BEGINS:)
```

PROV	SHRTNAME	CLASSDP	MINCS	PARSZCS	SOILTEMP	SOILMOIS
	1000 ROWS, NO HITS					
	2000 ROWS, NO HITS					
	3000 ROWS, NO HITS					
06	TACHE	O.HG	MIXED-CLAY	FNC	COLD	HUMID
06	TAILLON	E.DYB	MIXED-NONCLAY	FNL	COLD	HUMID
06	TILLY	GL.FHP	MIXED-NONCLAY	FNL	COOL	PERHUMID
06	TEMIS	O.G	NO-INFO	NO-INFO	NO-INFO	NO-INFO
06	TREMBLAY	O.HFP	MIXED-NONCLAY	COZ/SSK	COOL	HUMID
06	TERREBONNE	O.HG	MIXED-NONCLAY	FNL	MILD	SUBAQUIC
06	TADOUSSAC	E.DYB	MIXED-NONCLAY	S	NO-INFO	HUMID
06	TROTTIER	HU.LG	MIXED-NONCLAY	FNL	COOL	SUBAQUIC
06	TICOUAPE	GLSM.HFP	MIXED-NONCLAY	S	COOL	PERHUMID
06	TINGWICK	O.HG	MIXED-NONCLAY	S/L	COOL	SUBAQUIC
06	UPLANDS	O.HFP	MIXED-NONCLAY	S	MILD	SUBHUMID
06	VIGNEAULT	O.HFP	MIXED-NONCLAY	COL	COOL	HUMID
06	VIEN	GLSM.HFP	MIXED-NONCLAY	COL	COOL	PERHUMID
06	VILLEROY	O.FHP	MIXED-NONCLAY	S	COOL	HUMID
06	VALIN	O.HFP	MIXED-NONCLAY	S	COLD	HUMID
06	VAUDREUIL	O.HG	MIXED-NONCLAY	S/C	MILD	SUBAQUIC
06	VAUVERT	O.LG	MIXED-NONCLAY	COL	COLD	SUBAQUIC
06	WENDOVER	O.G	MIXED-CLAY	VFC	MILD	HUMID
06	WOODBRIDGE	GL.DYB	MIXED-NONCLAY	COL	COOL	HUMID
06	YAMASKA	GLSM.HFP	MIXED-NONCLAY	COL/C	COOL	HUMID

```
(RETRIEVAL COMPLETE)
ENTER 'END' OR NEXT REQUEST
?:
end
DO YOU WISH TO ACCESS ANY OTHER FILE ?
n
READY
logoff
A6230BK LOGGED OFF TSO AT 14:57:30 ON JANUARY 27, 1983
```

APPENDIX 6

CanSIS Output Using Standard Software Packages

Within a request the options for data selected, manipulated and output are literally infinite, Within the data submission procedures associated with a batch job, a limited number of options are possible; these must be specified. Basically these include:

(a) identification of the person submitting the job
(b) a routing address for delivery of the output
(c) JCL (Job Control Language) which describes the procedures to be run and the RAPID relation containing the data to be processed
(d) a series of selection criteria to extract a subset of data
(e) a number of data manipulation steps
(f) formatting and definition of the output including titles, report formats, column headings etc.

Two procedures for batch job submission are documented:

The first is a CLIST procedure LSISPMD whilch describes the steps for submitting and running one or a series of standard programs which have been developed within CanSIS. It has been developed for use with EASYTRIEVE programs written to access data from RAPID. LSISPMD provides the capability for specifying series of reports from one or severl RAPID relations. The selection criteria specified in LSISPMD must be appropriate for all the relations which are called. For the Soil Names file a standard report has been developed in EASYTRIEVE which outputs all the information in the file for selected soil name. It is invoked on TSO (time sharing option) from the 'ready' mode as illustrated in the following listing:

```
Establish contact with Datacrown

READY
ex sis01(lsispmd)    (This command tells the computer that you wish to submit
                      a standard report).

ENTER POSITIONAL PARAMETER SUBMITTER - (your surname)
lacelle
ENTER POSITIONAL PARAMETER ROOM      - (location where you can pick up
nb 30                                  your printout)

ENTER POSITIONAL PARAMETER PRIORITY   - (overnight, standard, weekend)
overnight
ENTER POSITIONAL PARAMETER APPLCODE   - (5100=Atlantic Prov., 5200=Quebec
5700                                     5300=Ontario, 5400=Manitoba
                                         5500=Saskatchewan, 5600=Alberta
                                         5700=B.C., 1000=General

ENTER POSITIONAL PARAMETER COPIES
1
DO YOU WANT OUTPUT ROUTED TO A TERMINAL - (if you want job scanned at terminal
n                                          answer "y", otherwise answer "n").
```

```
ENTER SELECTION CRITERIA IN QUOTES                    (Enter your province here)
OR RETURN TO END SELECTION
'if prov = 01'
ENTER SELECTION CRITERIA IN QUOTES
OR RETURN TO END SELECTION
(cr)
ENTER LIBRARY CONTAINING REQUEST OR END TO EXIT       (library containing
ag230.sis01.bmsrc.cntl                                 program)
ENTER MEMBERS OF THE LIBRARY CONTAINING REQUEST
OR ENTER END TO SEARCH FOR ANOTHER LIBRARY             (program)
names
AG230.SIS01.BMSRC.CNTL(NAMES)
JOB # (NAMES) submitted

READY
logoff
```

This report is ideally suited to generate reports for one or more provinces. The codes for province are identical to those used in the soil data (detail) file.

An example of the output of this report is listed in Appendix 6.1.

The second procedure for the submission of batch jobs is the general method of job submission in TSO. It requires that the job being submitted is completely defined including JCL, selection criteria, manipulation. Examples of this procedure are contained in Appendix 6.2.

The programs and reports described in this appendix illustrate special types of reports. They have been documented in detail as indendent units. Included in the write-up is a discussion of elements of the programs which must be changed, those parts which may be changed, as well as some suggestions as to the sorts of changes which can be conveniently made. These units provide examples of programs in Easytrieve and SAS and should assist users in developing their own specialized reporting and data manipulation procedures.

The example in appendix 6.3 illustrates normal batch job submission procedures. It is illustrated using the performance/management file but the principles apply to all CanSIS files on RAPID. It consists of:

> Rapid, Easytrive, SAS procedures - a description of the procedures for using Easytrieve to select data from a Rapid relation and write it onto a disk file and subsequently to define the data from the disk file into a SAS data base or library for statistical analysis.

SCAN

When an easytrieve job has finished running, the system will print out a completion message. If the output has been routed to a terminal the outpu is placed in a temporary file which can be scanned.

The procedure is as follows:

```
READY
scan names
LOG, JCL, SYSMSGS, 3 PRINT 0 PUNCH DATA SET (S)
```

The statement above lists the options the user has. You can evoke any of the following scan operations by typing in:

```
          gl log              (prints the log)
(or)      gl jcl              (prints the JCL)
(or)      gl sysmsgs          (prints any system messages)
(or)      gl print(1)         (prints data)
(or)      gl print(2)         (prints data)
(or)      gl print(3)         (prints data)
```

To get out of SCAN type:

```
          end                 (keeps data in temporary file)
(or)      end release         (prints whole job on the high speed printer
(or)      end purge           (deletes temporary file)
```

Appendix 7.

GUIDELINES FOR COMPLETING A REQUEST FOR PLOTS

(To be used in conjunction with the Spatial Display Point Plot Package User Manual prepared by CanSIS.)

The interactive TSO package DREAM allows a user to obtain tabular displays of the warden wildlife data on a remote TV screen or remote printer. These sheets will obtain a spatial display of the same data which can be overlayed to one of a variety of mapo projections. The plots are produced in Ottawa.

A. Description of data

- this should include the file or rapid relation from which the records are to be selected. How many records will be plotted?
- in addition do you want the data sorted or lumped in some way? (perhaps instead of plotting 10 different species, you want them lumped into carnivores, ungulates, etc.)

B. Data required

- give the date you require the plot

C. Description of use

- this will help CanSIS to understand the reasons for your specifications that follow. Will it be reduced to fit a report, coloured and framed for presentation purposes, or used for subsequent in-house analysis?

D. Plot

1. The size, colour and type of points are variable. Up to 14 different shapes and symbols can be used to reference different types of information. Labelling of these points is optional.

Information may be presented exactly as it exists in the data file, ie; labels may be actual numbers or names of observations, or they may be grouped into classes and plotted. Colours may be used to identify the group of symbols for example: red may be used for both x and + and green for both y and z; or, to identify classes within a symbol for example a red + may represent one to five observations and a blue + may represent 6 to 15 observations.

Note that the points are always plotted directly at the locations; there is no provision to offset different groups (eg. if red represents female and blue represents male they would be superimposed and indistinguishable in the plot).

a) The "symbols" are listed in the lefthand column.

b) The "type of information" is the value from the card;

eg.	+ griz	x blac
or	+ female griz	x male griz
or	+ spring/summer	x fall/winter
etc.		

c) If desired, a "label" can be associated with each observation, up to 30 characters in length and one space to the right of the symbol.

d) Enter "colour" (either blue, green or red) if other than black.

e) If there are any points that will be plotted outside the park boundaries but within the area you specified, enter the information in the space provided.

2 & 3 A legend and title are user specified. Three lines are allocated for the title. Size of letters is usually .25 inches and location is usually 1 inch below left hand corner of the plot.

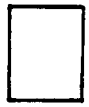
Up to 15 lines of legend can be specified. Size of letters are usually the same as labels, i.e. .10 inches.

Standard information that need not be included in title or legend

a) date and time
b) boundary map ID
c) plot ID
d) scale
e) number of points plotted

4. Virtually any scale can be selected. Realize the overall dimensions of the plot.

5. Specify the UTM or lat/long of the lower left-hand corners.

6. The map could be + + or + + 

7. Maps plotted on transparent materialks are suitable for overlaying. Maps on paper are more suitable for negatives and subsequent photo mechanical processes. Paper is also cheaper.

8. Ballpoint is cheaper and faster than fluid inks and fluid ink is cheaper and faster than scribing. Coloured ink is possible, but very slow (time consuming to clean ink reservoirs and pen nib).

9. Give the number of copies of the plot that you required.

E. Base maps are often essential to understand and interpret the plotted observations. Specify which separations you want; you can have one, two or three together. Screening base information will print it as a light grey; the plots will appear black.

REQUEST FOR A PLOT

Submitters Name _____ Ship plot to:

☐ same, or

Address _____ _____

_____ _____

Phone _____ _____

☐ Air express,

or _____

A. Describe the data you want to display.

B. Give the date you require the plot.

C. Briefly describe the use of this plot.

D. Plot

1. Specify types of information and symbol to be displayed for each; specify optional label.

Symbol	Type of Information	Label	Colour
+	_____	☐ blank, or _____	_____
x	_____	☐ blank, or _____	_____
□	_____	☐ blank, or _____	_____
○	_____	☐ blank, or _____	_____
△	_____	☐ blank, or _____	_____
◇	_____	☐ blank, or _____	_____
✳	_____	☐ blank, or _____	_____
⧖	_____	☐ blank, or _____	_____
⧗	_____	☐ blank, or _____	_____
Y	_____	☐ blank, or _____	_____
⌘	_____	☐ blank, or _____	_____
↑	_____	☐ blank, or _____	_____
⧖	_____	☐ blank, or _____	_____
|	_____	☐ blank, or _____	_____

Size of Symbols ☐.1 inches or, _____

Are there any points to be plotted that are outside the park boundaries but within the area you specified.

_____

_____

_____

2. Insert Title _____

_____

_____

Size of letters ☐.25 inches, or _____

Location ☐lower lefthand corner, or _____

3. Insert Legend

Legend Size _____

Symbol Explanation

_____

_____

_____

_____

_____

4. What scale is the plot? _____

5. The plot should be a square or a rectangle. Describe the four corners

of the plot of the basemap ☐same, or

_____ ☐ _____ _____ ☐ _____

_____ _____ _____ _____

6. Do you want only the corner reference points marked ☐ or the map boundary outlined ☐?

7. What material do you want the plot on?

☐ tracing paper ☐ stable base plastic (eg mylar)

☐ bond paper ☐ transparent overlay (eg. acetate)

8. What ink do you want? ☐ ball point ☐ fluid ink ☐ scribing

9. Number of copies of the plot required? _____

E. Basemap

What is the scale of the base map? _____ Do you want it photomechanically reduced ☐ or enlarged ☐ ?

Specify its I.D. _____

If a topo map, what features are desired?

☐ Rivers, roads, names ☐ Contours ☐ Utm grid

F. Do you want any manual enhancement?

	Yes	No
Do you want the plot and basemap on one sheet?	☐	☐
Do you want the topo information screened?	☐	☐
Do you want any shading or colouring?	☐	☐

G. Other instructions: _____

_____

_____

_____

_____

CANSIS STAFF ONLY

	Date
1. Test plot approved by author	_____
2. Final product sent to author	_____

APPENDIX 10

Generation of Flat Files from CanSIS FMS Files

The SELECT statement of Data Retrieval Language (DRL) is the main component of the flatfile procedure called AGSIS071. This allows one to select a subset of fields from any CanSIS file.

Select Statement

The SELECT statement has the following syntax:

```
SELECT element, element...
```

An element is defined as a field or segment. In the case of a segment all fields represented by it will automatically be generated. CanSIS allows repeating segments and therefore repeating fields. Only one SELECT statement is allowed per request. The easiest way to select a subset of recores is to generate a CanSIS-like subfile and then to apply the SELECT statement to the resultant file.

Procedure AGSIS071

This procedure converts all fields to one of: Quantitative (QTV), character (CHR), coded (COD) or Logical (LOG). Quantitative fields represent some real numeric quantitative value such as temperature. Character fields contain any combination of EBCDIC characters. Coded fields contain a code of a predefined set which represents a given characteristic. For example, codes'M' and'F' might be used to represent male and female for the field SEX. Logical fields contain a'1' to represent a true characteristics and '0' if it's false. When the byte alignmenbt option is used, they are zoned decimal 0 and 1.

Procedure AGSIS071 generates two reports which together describe the flat file produced. The first is the "Field Descriptor Report". This report describes each field of the fixed-length flat file records. A second report entitled "Flat File Codes Descriptor Report" lists all codes for coded fields mentioned in the Fields Report.

The JCL listing of procedure AGSIS071 further describes the options that can be used. These routines should be used only after consultation and assistance as required from the CanSIS systems project leader.

APPENDIX 11

Cartographic Data Structures

After a map has been degitized and the line and symbol data sets are merged, both data sets are then processed by editing programs that restructure the data into 4 files for storage and manipulation on the IBM/370 & 3033 computer. These files are colled Level 1, Level 2, Level 3 and Level 4.

Level 1 File

The Level 1 file contains an entry for each unique symbol on the map. Each entry contains flags, the actual symbol, addresses to point to the the first occurrence of that symbol in the Level 2 file, acreage of all areas with this identifier and the number of areas that carry this symbol on the map. This file is linked or connected to the Level 2 file by the address pointer.

Level 2 File

The Level 2 file contains an entry for each polygon (area) on the map. It contains symbol address pointers to Level 1 and Level 3 files, addresses of inclusions within the map area, next same symbol in Level 2 file, flags which indicate status of the symbol (deleted, linked, etc.) and polygon area.

Level 3 File

The Level 3 file contains the line coordinates. The actual x, y increments that give the line detail are stored in the Level 4 file. Each entry has address pointers to the Level 1, Level 2, the next entry in the Level 3 and the Level 4 file (with a value which is the number of increments). Since there is an area on either side of the line, information in the form of address pointers has to be stored for both the left hand side and right hand side of the line as defined by moving clockwise around the area.

Level 4 File

Level 4 is a massive array organized by a start position and the number of increments pointer in the Level 3 file. In the Level 4 file data can be manipulated independently from the other cartographic data. In most cases the Level 4 file is interrogated only for plotting purposes.

Editing Carographic Data

The data files coming from the digitizing table are run through the line and symbol editor programs. Redundant data are removed, angle and area calculations made and the data is stored in the various files.

Synbol Address Calculation

Before the computer begins to build the map, the location of the data set for each symbol has to be determined. This is done by the matrix address generation algorithm which greatly facilitates map linkage and map retrieval. An imaginary 1" grid if placed over the map. Each symbol is recorded within a grid cell by its coordinates in the primary x, y grid matrix. Since it is not uncommon to have more than one symbol in a 1" square, a secondary matrix is used to record up to 3 entries in the primary matrix. If more than 3 entries occur, an address is generated to the next free element in the secondary matrix for 3 more symbols.

Line Address Calculation

The location of each line can be calculated through the matrix address generation algorithm as well. In this case, an imaginary 1/4" grid is superimposed on the map. In the primary matrix, an address to an entry in the Level 3 address array is given in the grid cell where the start and the end of the line are located. AGain up to 3 line entries can be recorded in the Level 3 secondary matrix array for each 1/4" cell. If more entries are required, the fourth location will contain an address pointing to the next free location in the secondary matrix. Since the start and end of a chain normally fall in different grid cells, each line has two entries in the Level 3 address arrary. The line detail is stored in Level 4. The address and the number of points that comprise the chain are stored in the chain's record in Level 3.

Establishing Levels

Through linking, the relationslhip is established between the symbol identifiers and the line chains. The program reads sequentially through the Level 2 file adding to Level 1 each unique identifier found and building the links as it goes. The procedures is as follows:

The first and last of coordinates that define one line chain are used to locate that line in the Level 3 file using the matrix addresses. Since several line chains may start or end in a given 1/4" square, a match has to be done on the basis of both the start and end coordinates. In the event that two or mor chains have identical start and end coordinates, a third coordinate (on the line) in conjunction with the line detail in Level 4 is used to find the correct line. When a match is made, the pointer to the first line chain is entered in Level 2. The end of the given line now identifies the intersection into whlch the next line runs. All the lines coming into this junction are located by the matrix addresses. The angles of the lines are determined and the line with the minimum angle is chosen.

A number of checks are done at this point to assure that the chosen line is not already linked up to another polygon, and to establish whether the given polygon has now been completely enclosed. In addition, the unit identifier is assigned to the proper side of the line, the pointer to the last line is established and the area under the curve is added to the acreage total for the polygon. After the parent polygons have been linked up, the island or inclusions are linked to the parent areas with the total area adjusted accordingly.

The Level 3 file structure is very efficient. Each record contains all the information required for linking 50 line chains. Since digitizing is conducted in an orderly fashion, I/O acitivity is very low during the linking process, as most of the line chains for a polygon are in the same record in Level 3.

If the program has problems linking an area, it will unlink the area being worked and the area involved in the violation. An error message is generated and both the error and its locatin is stored in an erro file. This is used at the end of the run to produce an error plot.

Audit Trail

A black and red color plot and printout is generated for error correction. Everything that is plotted in black is properly linked, red signifies errors. The errors are displayed on the plot but detailed on the printout. Common errors are missing lines, symbols, unidentified intersections, misspelled symbols, and incorrect identification of the first line. These all generate different error messages so the problems are quickly diagnosed. A new data set is created containing the corrections whlch is merged with the original. Typically 3 or 4 updates will produce a clean map. Unusual problems may be more difficult to correct. After the first or second pass, the map is also edited cartographically. Any updates resulting from this operation are merged with the next update.

Appendix 13.2

Agriculture Canada

REQUEST FOR INTERPRETIVE MAP FOR COMPUTERIZED LEGENDS

DEMANDE DE CARTE INTERPRETATIVE POUR LES LEGENDES INFORMATISEES

Submitter's name and address - Nom et adresse du demandeur:

Ship map to - Expédiez la carte à:
☐ Submitter
☐ Demandeur

Phone - Telephone:

Ship by: / Expédiée par: ☐ Air Express ☐ Other / Autre _____

1. Library containing Easytrieve Program - Bibliothèque contenant le Programme "Easytrieve":

2. Members - Membres:

3. List map Easytrieve program is to be run against - Indiquer les cartes devant être traités par le programme Easytrieve:

CanSIS MAP ID #1	CanSIS MAP ID #2	CanSIS MAP ID #3	CanSIS MAP ID #4

4. Describe interpretive map to be retrieved - Décrire la carte interpretative voulue:

5. Date required - Demandée pour le:

6. Symbol size - Taille des symboles:
☐ A-10/100" ☐ B-8/100" ☐ C-6/100"

7. Title - Titre:
Size of letter - Taille des lettres:
☐ .25" or/ou: _____
Location
☐ Left hand corner / Coin gauche inférieur
or/ou: _____

8. Interpretive Map Scale - Echelle de la carte interprétative:
☐ Original Plot / Plan original ☐ Other: / Autre: _____

9. Is a preliminary copy required / Désirez-vous une copie préliminaire
☐ Yes / Oui ☐ No / Non

10. To assist cartography in producing a final map suitable for your needs please answer the following questions. Additional assistance with items # 10-16 can be obtained by contacting the Chief Cartographer, LRRI at: (613) 995-5011 (220)

Pour aider le service de cartographie à produire une carte finale qui répond à vos besoins, prière de répondre aux questions suivante. Pour de plus amples renseignements sur les questions 10 à 16, communiquer avec le cartographe en chef de l'institut de recherches sur les terres, au (613) 995-5011 poste (220)

i) Indicate use of final copy - Précisez à quelle fin servira la copie finale

ii) Will additional copies be reproduced - Des copies additionnelles à reproduire:
☐ Yes / Oui ☐ No / Non

iii) Will copy be included in a report / La copie demandée sera-t'elle inserée dans un rapport
☐ Yes / Oui ☐ No / Non
If yes will it be used / Si oui, sera-t'elle utilisée ☐ same scale / même échelle ☐ reduced / réduite

iv) Do you require a photomechanically produced clear base positive / Voulez vous une epreuve positive transparente photomécaniquement
☐ Yes / Oui ☐ No / Non
If yes allow 5 working days for reproduction / Si oui prévoir 5 jours ouvrables pour la production

11. Ink - Encre:
☐ Ballpoint / Stylo à bille ☐ Fluid Ink / Encre liquide

12. Plotted on - Support:
☐ Tracing paper / Papier à calquer ☐ .004 matte surf. film / Film de surf. matte 004 ☐ .007 matte surf. film / Film surf. matte .007

13. Pen colour - Couleur de l'encre:
☐ Black / Noire ☐ Blue / Bleu ☐ Red / Rouge ☐ Green / Vert

14. Scribing - Tracé sur couche:
☐ Yes / Oui

15. Lineweights will be chosen by cartography unit. Indicate if any data requires special treatment, i.e. heavy symbols or heavy polygon boundaries, etc.

Épaisseur du trait sera choisie par le service de la cartographie. Préciser si certaines données requièrent un traitement spécial (symboles ou contours des unités pédologiques en caractères gras, etc.)

16. Number of copies required: please circle number
Nombre de copies requises: prière d'encercler le nombre voulue: 1 2 3 4 5 6 Other/Autre: _____

17. Is a base map required / Désirez-vous une carte de base ☐ Yes / Oui ☐ No / Non

If yes, scale of base map / Si oui, echelle de la carte de base ☐ Original scale / Échelle originale ☐ or/ou: _____

MAP ID:

If a topo map, what features are desired / Si c'est une carte topographique: caractéristique ☐ Rivers, Roads, Toponomy / Rivières, routes, toponomie

☐ UTM Grid / Grillage U.T.M. ☐ Contours / Courbes de niveau

18. Do you want any manual enhancement / Désirez-vous quelques retouches manuelles ☐ Yes / Oui ☐ No / Non

If yes: / Si oui:

Plot and base map on one sheet: / Plan et carte de base sur une même feuille: ☐ Yes / Oui ☐ No / Non

Topographic Information screened: / Renseignements topographiques atténués: ☐ Yes / Oui ☐ No / Non

Shading or colouring: / Teinte ou couleur: ☐ Yes / Oui ☐ No / Non

19. Special instructions - Instructions spéciales:

20. Unless otherwise stipulated the plot tape will be released after the final copy of your plot has been sent. If a longer retention period or more plots required, please indicate the date the tape can be released.

Sauf indications contraires, le ruban du plan sera extrait du système après la livraison de la dernière copie de votre plan. Si vous désirez une plus longue période de retention ou plus de plans, préciser la date à laquelle le ruban peut-être extrait

Date _____

Canada

AGR 2369 (83 08)