Uses of Soil Information Systems

Proceedings of the Australian Meeting of the ISSS Working Group on Soil Information Systems Canberra, Australia, March 2-4, 1976

Alan W. Moore and Stein W. Bie (Eds)

Centre for Agricultural Publishing and Documentation
Wageningen, the Netherlands
1977

ISBN 90-220-0638-7

Cover design: Pudoc, Wageningen

Printed in the Netherlands

Contents

Note on currency exchange rates

In some papers costs and benefits are expressed in Australian dollars, 1970-1976 values.

International exchange rates are unstable at the time of this book going to press. As indications the following approximate rates were effective per 1 April 1977:

A$ 1.00 = US$ 1.10 = Dutch Guilder 2.50

The editors

Foreword

E.G. Hallsworth, Land Resources Laboratories, CSIRO, Adelaide, South Australia

The Tenth International Congress of Soil Science, held in Moscow in August 1974, provided a point of contact for a number of people interested in the use of computers in soil information systems. Prior to that time the Soil Resources, Development and Conservation Service of FAO had attempted to foster some interest in this area in several ways. This proved less feasible than hoped because of the limited resources available and the inadequate computing facilities that were available in Rome in the early 1970'ies.

A meeting was held in Rome in April 1972 under the auspices of FAO and Unesco, which brought together scientists from Europe and North America to exchange ideas. The initiative for fostering communication is now shared with the International Soil Science Society. At the Moscow congress, in Commission V, it was agreed to set up a Working Group on Soil Information Systems, with the objective of encouraging the free exchange of ideas and experi ences among soil scientists interested in automated data handling, particularly as this often requires large investments in expertise and equipment.

The Working Group was formally constituted during the first meeting in Wageningen in September 1975, with Dr Ir J. Schelling and Dr S.W. Bie of the Netherlands Soil Survey Institute as chairman and secretary respectively. Over 50 participants, mostly from Europe and North-America, were present at this meeting which included four days of papers and discussion plus a one-day field excursion.

It was known that there was a number of soil scientists in Australia also interested in these developments. It was felt that it would be worthwhile to hold a second meeting of a similar nature in Australia. Originally it was hoped that the two meetings would be held more or less concurrently, but the lead-time necessary for organizing the meeting in Australia resulted in it eventuating some six months after that held in Wageningen. We had the advantage of our European and American colleagues in one sense, as we had available to us the proceedings of their meeting which were published in November 1975.

The Australian meeting was somewhat different in that the number of participants was smaller, giving a group of a size to allow a free flow of informal discussion. Its main value was seen as the building of initial contacts between actively working or inter-

ested in the area of information systems for soil and related data. It was not intended that this meeting discuss data banks per se, buth rather experiences in the use of various techniques associated with information systems, for the most part computer-based. It dealt mostly not with what data should be collected or for what purposes (the province of the soil scientist alone) but whether and how we can collect and process data with computer assistance to meet defined objectives more efficiently than by manual handling (the province of the computer scientist in consultation with the soil scientist). This is not to suggest that the second is more important than the first, but merely emphasizing that it was the one chosen as the topic for this meeting.

It can be seen from the papers presented at this workshop, and published herein, that many of the early technical difficulties encountered in the development of machine-based soil information systems have now been overcome. The main problems that remain are for the most part human ones - the difficulties of persuading administrators of the value of the systems, yet at the same time avoiding the pitfall of expecting too much of them. Difficulties are also likely to be experienced at the input end, with resistance by surveyors and other soil scientists to the use of an unfamiliar, highly technical system. The greatest problems seem to be those of compatibility, not only in the methods of storing and retrieving the data, but also between the scientist concerned with collection of information, data base manager and user. The discussions at this meeting highlighted the difficulties likely to be encountered if a system becomes too complex, or attempts to cover too wide an area.

They also highlighted the great good that can come of putting together a group of people, actively concerned with the problems, and allowing them to talk freely of the ways in which they may be solved. This, surely, is the main role of the International Society of Soil Science.

Finally, it is with pleasure that I acknowledge financial assistance towards publication of these proceedings provided by the following organizations: Federal Council of the Australian Soil Science Society Inc; International Soil Science Society; N.S.W. Branch of the Australian Soil Science Society; New Zealand Soil Science Society; New Zealand Ministry of Works; New Zealand Soil Bureau; Consolidated Fertilizers Ltd; Departments of Agriculture of Western Australia, South Australia, Victoria and New South Wales; Soil Conservation Service of New South Wales; Victorian Soil Conservation Authority; and Commonwealth Scientific and Industrial Research Organization.

The Canberra Meeting reviewed

A.W. Moore, Division of Soils, CSIRO, Brisbane, Queensland
N.M. Dawson, Department of Primary Industries, Brisbane, Queensland

Introduction

The meeting reported in this book brought together most scientists in Australia and nearby countries interested in the area of information systems for soil and related data. The 25 participants represented all states of Australia plus a leavening of two New Zealanders and one Malaysian. This relatively small number allowed an informal approach and this helped to contribute to the success of the workshop.

In a local context, the meeting was timely in view of the discussions currently taking place in Australia on the feasibility of some sort of national approach to soil survey. Although this was not on the agenda for this meeting, there appeared to be a consensus of opinion among participants that the time was not ripe for the establishment of a national (centralized) soil data bank. However, the possibility of establishing regional data banks (at state or more local levels) was considered to be worthy of examination.

For a given soil information system, whatever its size or scope, internal compatibility of data is essential. In the case of those systems discussed at the workshop this requirement had been met, though not without difficulty in some cases. The number of people involved in the various information systems varied from one or two to many; in general, the more people involved the more difficult it is to achieve compatibility.

By design, the workshop did not consider the question of compatibility of data. This is a topic largely separate from that of data base management techniques, with political and administrative overtones as well as technical aspects. At anything other than a local, individual organization level, the problems of compatibility would have to be resolved by agreement on a national scale, if the evolution of widely-used integrated data bases were an ultimate objective.

The workshop covered a range of computer usage from simple to sophisticated, of size of data base from a few tens of thousands of characters to several millions, and a spectrum of data base applications from specialized soil test and soil micromorphological data banks to broad land resource surveys.

Unpublished contribution to the meeting

Eleven of the papers given at the workshop are published in this volume. In addition, two other papers were presented but have not been published here for various reasons. For completeness, a brief indication of their contents is given below.

R. Lee (Computer processing of soil survey data: a case study from the West Coast, South Island, New Zealand) discussed the use of a system devised in the N.Z. Soil Bureau for processing soil survey data. A detailed discussion of field sheets and data processing is published elsewhere (Lee et al. 1976).

Data were punched straight from field cards onto paper tape and entered into the computer system. After checking and editing the following types of output were produced:
1. various indexes to the data,
2. text profile descriptions corresponding to the coded descriptions on the cards, and
3. frequency distribution listings showing the number of times within a taxonomic unit that particular horizons had been described and the number of times within these horizons that each descriptive term had been used.

From the frequency distribution listings, information was extracted on modal characteristics and variability of the major soil profile features within each taxonomic unit. The computer used was an Elliott 503 and all programs were written in ALGOL. A total of 1100 profile descriptions were processed through the computer system during the course of the surveys.

In comparison with manual methods, the system reduced the time taken in data preparation for survey reports by up to one third. Cards of uniform design also helped to achieve a consistent standard of profile description and reduced the time required to make such a description.

R.K. Rowe, J.N. Rowan and W. Papst (Using soil and land resource data for soil conservation in Victoria) outlined the magnitude and complexity of the land resources data handling requirements of the Victorian Soil Conservation Authority which has led them to the conclusion that a computer data base is desirable.

Systematic land system surveys currently cover about 60 per cent of the State. Because of increasing demand for this kind of information, the rest of the State was mapped in 1975 using air-photo interpretation with very limited field checking. Information from the tow levels of survey has been combined in a provisional State Land System Map on the 2-miles-per-inch County Plan series. Land system and component data are presented in 540 land system diagrams, which would average about 3 components each. Thus some 1600 components are recorded with different assemblages of climate, parent material, vegetation, topography and soils.

A trial storage and retrieval exercise has been carried out with a small set of land system dat using INFOL on the CSIRO CYBER 76. Difficulties experienced with INFOL have caused them to consider developing a system based on FORDATA (Smith and Mackenzie 1976). This will store land system, component and site data, with the latter stored according to system and component. Spatial representation is envisaged using references to the national mapping grid.

It is not intended at present to computerise the mapping unit boundaries. These will be retained in a library of 1:100 000 topographic maps of the Australian Map Series or on photo-indices at the same scale based on the same map grid. The location of each mapped area will be "addressed" by the grid reference of a point within the area, and the addresses of all the mapped areas which constitute a land system will be included in the data set for that land system.

Demonstrations of operations on existing data bases

The final day of the workshop was devoted to practical demonstrations of the use of two existing information systems. These were the computer data bases associated with (1) the South Coast Project (Cook 1975) of the CSIRO Division of Land Use Research (LUR), and (2) the Western Arid Region Land Use Survey (WARLUS) and the Upper Burdekin Catchment Survey (UBCS) of the Queensland Department of Primary Industries.

In brief, the former consists of (1) a data base containing boundaries of approximately 4000 maps faces which are based on cadastral and physical land units, and (2) an attribute data base for which the FORDATA data base management system (Smith and Mackenzie 1976) is used. These two data bases are linked by cross-addressing.

As an exercise, the following request was proposed by a participant: Where can improved pastures be developed in an area extending westwards from Bateman's Bay to the mountain divide? (This is included in the area covered by the South Coast Project.) It was of interest that it took approximately half an hour to formulate this request in terms of areal and point data stored in the data bases. Criteria eventually used included slope, landform type, relief, rock outcrops, stone size, vehicle mobility rating, soil depth and land development status. Two applications programmers coded the request within half an hour and the required map (with 5 classes of suitability for improved pastures indicated) was output on the plotter within two hours.

The second demonstration consisted of operating on the latter two data bases mentioned above, via a terminal, using the methods described by Cormack (1976). On the WARLUS data base (point data) the following was done: (1) extraction of morphological data for the top three recorded layers in the soil profile, (2) extraction

of laboratory pH and electrical conductivity measurements and sorting on soil group and profile class, (3) presentation of frequencies of electrical conductivity classes for individual soil groups and of pH classes for profile classes, and (4) retrieval of a set of data in response to a request formulated by participants. From the UBCS data base (grid cell data) the following were produced: (1) listings of mapping units and areas which satisfy specified pastoral capability criteria (giving rise to 5 classes) within the Conjuboy 1:100 000 map sheet, and (2) intensity maps of the pastoral capabilities for the same sheet.

The commands and output for these operations (with the exception of (4) above) are shown on the microfiche stored in the pocket inside the back cover of this publication.

Most participants felt that these execrises provided a useful complement to the papers presented during the previous two days, giving them a clearer insight into the advantages and problems of two different approaches to information systems through actual examples of data manipulation by means of working data base management systems.

Overview of the meeting

General comments

The need for good computing facilities, easily accessible, became apparent, although it was cheering to observe that useful work is being done in spite of poor facilities. Lee (New Zealand) had to employ an aging Elliott with poor edit software and no interactive facilities. Stackhouse (Tasmania) outlined the drawbacks he encountered in being located 300 km from the nearest computing facility that he could use.
Scientific staff need to have ready access to computing facilities (through terminals located in their place of work if necessary) and, perhaps more important, to computing professionals. There was a diversity of computers (e.g. Elliott 503, CDC Cyber 76, CDC 3200, ICL 1901A), languages (ALGOL, FORTRAN, COBOL) and data base management systems in use among the workshop participants. While not of great significance for the particular tasks that each set out to do, this has obvious implications for any wider cooperative projects. For example, if an integrated data bank for a national soil survey were envisaged it almost certainly would have to be based on CSIRO's computing network, CSIRONET, because of its nation-wide coverage and the fact that in Australia major developments in data base management applied to soil science have been related to it.

A fact well known to computer scientists but perhaps less well recognized by soil (and other) scientists is that it is not possible to develop a data base management system tailored to a small data bank and simply scale it up for use with a large data bank. This is very evident from the considerations of data structure

(internal schema) presented in the paper of Mackenzie and Smith.

Input

Discussion of techniques available for collection and input of field data to the computer (Lynch's paper and others) suggested that some form of hard copy recording in the field with subsequent transcription to computer-compatible input (punch cards or paper tape in most instances) is still the most practical way of doing the job at present. The use of mark-sensing, OCR, field recording on tape, etc. is largely impractical because of the difficult conditions usually encountered in field survey and the need for easy visual checking of records. Line-follower digitizing is almost the universal method of inputting boundaries (for areal data) to the computer; a difference of opinion as to whether it is preferable for this to be interactive or not was not resolved.

Storage and retrieval

Eminent good sense was evident in the choice of appropriate methods for handling the data banks with which the participants have been working. The data banks ranged from a simple two-dimensional array of numbers, which could be held in memory in the Cyber 76 (Colwell), to a very large complex data bank of land resource data (Cook and Cormack), handled by FORDATA (Mackenzie and Smith) and a specially developed system for storage of digitized boundaries. Although there appeared to be some divergence of opinion on the merits of the grid-cell and boundary (or polygon) approaches to storage of areal data, discussion brought out the fact that the basic philosophy behind them was the same in the two instances. Users of either approach have to accept that, no matter at what scale data is recorded, heterogeneity exists within map units and map faces. Most participants seem to have accepted the need for overlaying maps of various attributes and thematic maps outside the computer (i.e. before input) rather than inside.

There was some discussion on the desirability of the development of query languages (non-procedural languages) in association with storage and retrieval systems. On this hinges to some extent whether the user had direct access to the data base or needs to access it via an applications programmer. There was a divergence of philosophy on this; for example, user operation is aimed at by the Queensland Department of Primary Industries (Cormack) while applications programmers are needed to make use of the data base management system for the South Coast Project of the CSIRO Division of Land Use Research (Cook).

Display

While display may appear a trivial consideration, this is by no means true. Reference is made below to the need to be able to

present information in familiar hard copy form if initial resistance by users to computer assistance is to be overcome.

For map display both line printer (Armstrong and Wetherby, Cormarck) and line plotter (Cook) output was demonstrated. Choice of one or other approach is dictated by the projected use of the output and by cost. The former is less accurate and generally less visually acceptable but much cheaper. The potential usefulness of non-impact plotters for displaying grid-cell data was mentioned.

For tabular and textual material the line printer is used almost exclusively at present. The desirability of having comprehensive report generators associated with storage and retrieval systems was proposed by one participant. It was felt that recent technological advances have made this an area where rapid change can be expected. For example, facilities such as COM (Computer Output Microfilm) could lead to greater efficiency by eliminating errors arising during transcription steps in the preparation of material for publication, and with microfiches as output appear to provide a means of publishing large amounts of raw and processed data at very low cost and with little bulk added to a publication.

People problems

The human factors involved in changing from manual to computer handling of data were mentioned by a number of speakers. It is difficult to get persons at the managerial level educated as to the possible uses of a computer and, once a decision has been made to use it, there is often a problem of resistance of field workers to changed methods of field recording, etc. The latter usually recognize the usefulness of the new approach once the initial resistance is overcome. At the other end several people commented on the need to be able to extract data easily and to present it in a form familiar to the user if he is to accept the new procedures. In small systems, however, the data collector and the user are often the same person. The interface between human being and computer is also important. For example, card punch operators must find the punch documents acceptable or nothing gets into the computer.

A frequently made point was the desirability of close cooperation, between providers and users of information on the one hand and computer scientists on the other, in the design and operation of information systems. The expertise of computer scientists is essential when large, complex data bases are involved. There is a continuing need to inform potential users of the availability of systems and/or data that may be of use to them. The workshop has made a start as far as systems are concerned.

Costs

While it would seem to be axiomatic that computer-oriented data

base management system should be competitive with alternative systems, in practice it is difficult to cost these systems accurately (see, e.g. Dawson and Moore). These difficulties are compounded by marked differences in charge rates between different computer installations and between different classes of user for any given installation. However, most of those who had used a computer as an aid in their own particular work felt that its use improved their efficiency. This was so even if there was no actual "economic" benefit - often in this situation money was being substituted for the time of skilled staff. In the long run, time is on the side of the computer as costs of computing have risen more slowly than labour costs and this trend shows no sign of changing.

Conclusions

1. The meeting achieved its aim of bringing together most of those interested in information systems for soil and related data. There is a need for continuing contact between participants who are working actively in the field.

2. It appears from the papers and discussions at the workshop that sufficient experience now excists to overcome most of the technical problems which might be encountered.

3. Although difficult to evaluate, it is likely that costs associated with computer-assisted soil information systems are of the same order as those handled manually, but the former have the advantages of increasing the time available to scientists for professional work and of allowing data processing and analysis not previously feasible.

4. The use of microfiche appears to offer a suitable means of publishing the large amounts of data that accumulate during resource surveys. Bulkiness and cost of printing have mitigated against such publication in the past. Where this data is held in computer-readable form it can be output via a COM unit with very little extra work or cost.

5. Because of its wide coverage of Australia and the present use of it by the two groups most active in the area of information systems for soil and related data, CSIRONET appears to be the logical choice of a computing facility for such enterprises in the foreseeable future.

References

Cook, B.G. (1975). A computer data banks in a regional land use study. In URPIS-THREE: Proceedings of the Third Australian Conference on Urban and Regional Planning Information Systems, Newcastle.

Cormack, R.S. (1976). Data delineation and computer techniques for line printer mapping and tabulation. In these Proceedings, p. 58.

Lee, R., M.J. Newman and A.R. Gibson (1976). Computer processing of soil profile data from surveys in New Zealand. Geoderma 16: 201-209.

Smith, J.L. and H.G. Mackenzie (1976). FORDATA: A data base management package under Fortran on the CDC Cyber computers. CSIRO Aust. Div. Computing Research, Canberra

Input methods and facilities available for land survey data

L.G. Lynch, Soil Conservation Service of N.S.W., Sydney, New South Wales

Abstract

A survey has been carried out by the Soil Conservation Service of N.S.W. on the methods and facilities available for the input of land information data to a computer system. The sources and methods available for input are described, with their advantages and limitations discussed. An outline of the types of input facilities available, particularly digitizers, is given. Features which should be considered in the selection of a suitable configuration are discussed.

Introduction

Any information system for soils and/or related data presents three aspects:
- data collection and input,
- data storage and retrieval,
- application, including display.

This holds for both manual (Beckett et al 1972) and computer-oriented data banks.

In order to assess the feasibility of any data bank, all three aspects must be considered. Some people have considered the retrieval and display aspects of the system only. In these fields, a computer-oriented system usually has an advantage over the corresponding manual system. However, the advantages of using a computer for input and storage are not so obvious. Depending on the sources of data, the computer system may be more costly than the manual system for input. Maintaining the storage file, particularly with regard to updating, may also be more costly. Although the computer system may not be economically feasible in the short term, it may be of enormous value in its speed of retrieval and display once the data base has been established.

In designing a computer-oriented system, the following facets of data input should be considered:
- it is usually the most time-consuming aspect,
- it must be related to the sources of input and the capabilities for storage and retrieval on the computer,
- it must also be related to the end-product required from the system,
- it can be restricted by the availability of facilities for input.

Sources and methods of data input

There are four main sources of input to a data base:
- data collected in the field,
- aerial photography with field checking,
- existing thematic and contour maps,
- satellite imagery.

Data collected in the field

This source is generally applicable to soil surveys where much of the information can only be obtained from field observations (i.e. profile data, etc.). Wetherby and Armstrong (1975, pers. comm.) and Webster and Burrough (1972) have completed surveys in this manner. A regular grid is placed over the study area. At each grid point, attributes are measured and usually recorded in coded form onto data sheets for transfer to punch cards, tape or disk.

Aerial photography with field checking

This is in Australia the most common method of obtaining data for an information system. Many attributes can be assessed by this method. The Soil Conservation Service of New South Wales has measured topography and slope, existing land use, erosion, vegetation, geology, soil, landform and drainage patterns by this technique.

Traditionally the presentation of land resource information has been in the form of irregular land units whose description is accepted as being internally homogeneous. However, with the variability of land, this homogeneity is unattainable in pratice. The differentiation of land units thus becomes a subjective interpretation of boundaries relative to their importance.
In preparing data for an information system, this apporach could be used with single or multi-attribute thematic maps (the input of this information into a system will be discussed in the next section).
Another approach is to overlay a regular grid over the study area. A description of the land can be made for the cells within the grid mesh by:
- point sampling - the land at the centroid of the cell is described,
- model sampling - the predominant land unit for the cell is described,
- probability sampling - the cell is described by a number of units with a weighting given for each unit based on the proportion of the cell it occupies.

With these methods, land units need not be delineated. They are sampling techniques. There are techniques for minimising the sampling error.
- By using probability sampling in preference to modal and point sampling. (Similarly modal is preferred to point sampling.)
- By using a smaller grid mesh and cell size. The degree of rep-

resentation that will be obtained of the area will be greater with a smaller cell size (Nichols 1975). If critical areas (land slip, mining, severe erosion, drainage lines etc.) which are generally very localised in occurrence need to be represented, a relatively small cell size may be necessay. In many studies (Cormack and James 1975, pers. comm.; Venz 1975, pers. comm.), the size of the cell has been determined by the final maps produced by computer line printer. This need not be the case. Hoeske and Wilson (1974) use a graphics plotter for producing maps at any desired scale. However, plotting by this method is very slow. Cram (1975) has used a matrix plotter. This device gives the flexibility and speed required. Therefore, the cell size should not be limited by output device. However, it may be limited by the resources for storage on the computer.

- By using variable cell size. In areas of particular interest or where the land is more variable than the remainder of the study area, small cell sizes can be used. Hoeske and Wilson (1974) have used this approach.

In order to illustrate some of these points, a small sample area was taken where land units based on one attribute had been delineated (see Fig. 1). Three sampling methods were used over the area.

- Grid mesh with point sampling.
- Grid mesh with model sampling.
- Smaller grid mesh (four times as many cells) with model sampling.

The discrepancies with the cell sampling to the land units in classifying the area were determined (see Figs 2, 3 and 4). The extent of these discrepancies were calculated as percentages of the total area. These were 42, 27 and 16 percent respectively. This demonstrates that a better representation of the area is obtained with model sampling (in preference to point sampling) and smaller cell sizes. Actually the smaller cell size with modal sampling gives a good representation of the area. The attribute boundaries described by the cell method would be well within the tolerance of the subjective assessment of land unit boundaries. The input of grid data to be a computer information system may be done manually. The attribute data for each cell is coded onto data sheets by the interpreter. Field checking of the data should be done after this stage. The information is then punched directly from the data sheets.

Existing thematic and contour maps

In many cases, existing thematic maps (e.g. soils, vegetation, landform) are available for the study area. Contour maps, from which aspect, slope and, to a certain degree, terrain can be determined, may alos be available. These data may be stored in two forms:

- as the unit boundaries as defined (Cook and Johnson 1973; Cook 1975),
- as cellular data (Cram 1975).

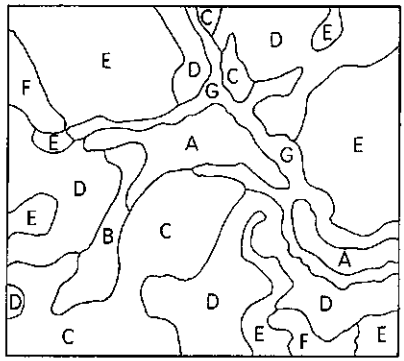


Fig. 1 Sample area

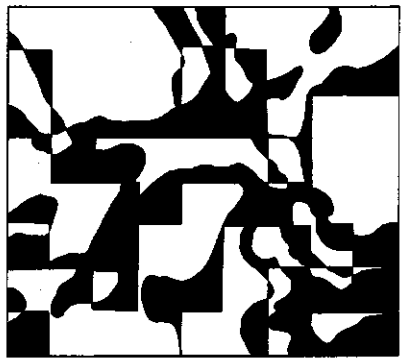


Fig. 2 Discrepancies with point sampling

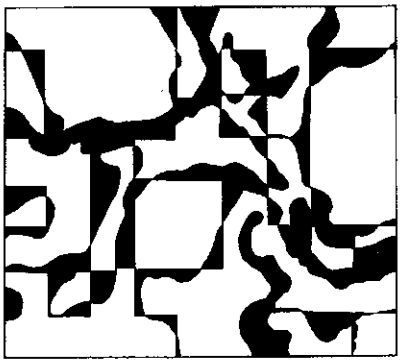


Fig. 3 Discrepancies with modal sampling

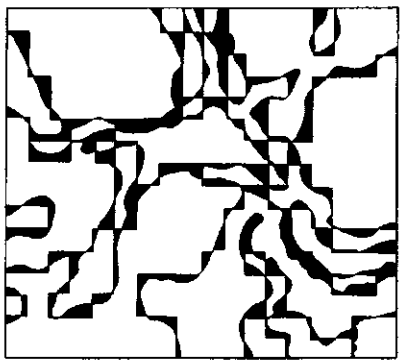


Fig. 4 Discrepancies with modal sampling and smaller grid mesh

The method of storage has various implications on the method of data input.

- If a cellular storage system is used, a manual method of input (as described for aerial photographs) can be used. On the other hand, a digitizer must be used with a boundary definition system.
- The accuracy required for cellular data storage is not as important as with the definition of boundaries where use of a digitizer is necessary. With the latter system. overshoots and boundary mismatches create problems.
- With the boundary definition system, the overlaying of attribute maps and requires editing to remove minute regions and near-coincident lines. This may be done on the computer or cartographically by preparing a attribute complex map. Due to computational diffi-

culties Cook and Johnston (1973) and Cook (1975) prefer the cartographic method.
- The cellular system is more open-ended than the boundary system particularly if the attribute complex map has to be again prepared for the latter system to incorporate further attributes.

Another method of incorporating this type of information into a data base, is to use an irregular cell basis. Humphries (1973) has described a system where the basic cells are determined by one attribute (allotments from cadastral maps). These subdivisions are located by defining the corners of the allotmens. For each allotment, attributes such as land tenure, soils, land use, etc. are described. The degree of representation obtained with this method is determined, as with grid sampling, by the size of the cell.

Satellite imagery

Satellite imagery has offered the photo interpreter a substantial expansion of source material. A description of satellite images is given by Emery (1975). The usual format of the images is black and white prints produced at a scale of 1:1 000 000. The image in digital form is stored on magnetic tape.
Automatic interpretation of the satellite images involves the clustering of the spatial elements into similar units related to ground data. It has two major limitations.
- It is a very expensive process. Even elementary enquiries are costly due to the enormous amount of information stored in digital form on the magnetic tapes.
- Very complex algorithms are required to detect different land untis. These algorithms would have to be based on criteria used by the aerial photograph interpreter for differentiating units (as described earlier in this paper).
However, Emery (1975) states that there are advantages of satellite imagery in land resource investigations:
- the broad synoptic view obtained in one scene which in his case covered an area of approximately 34 225 km^2,
- it permits the initial delineation of resource boundaries which can be supplemented at a later date by large scale photographic interpretation or by field survey,
- the availability of multi-date imagery.

He lists the disadvantages as:
- The scale of the imagery.
- Cloud cover.
- Lack of stereo viewing.
- In Australia, the lack of a receiving station to acquire real-time imagery.
- The lack of computer hardware in Australia suitable for the automatic processing of the imagery. This will be discussed later in the paper.

Choice and definition of attributes

Potential users of a computer-oriented information system should be aware of the capabilities of a computer for storing and manipulating data. However, they should not exaggerate the potential of a computer system simply because it has the capacity to store and manipulate more attribute data than a manual system. There should be some rationalisation with regard to the number of attributes being measured and the classification within them, based on the aims of the study. If the study is a general survey of the area, there would normally be less attributes and less definition than if the study is for research purposes, where simulation and testing of hypotheses and comtemplated.

Validation of data

An important aspect of data input to any information system is editing. It is fairly easy to edit locational data by producing an image or map of the area. However, other data is more difficult to check. Attributes can be checked for mutual exclusiveness (e.g. flood plains are not likely to occur on slopes exceeding 20%). Maps of certain attributes, or combinations of them, may be prepared and field checked. However, when ons has 20 000 cells or functional units with 10 attributes, this is an enormous task.

Facilities for data input

Manual methods of data input can be used for data collected in the field land for aerial photograph interpretation where a regular grid is used. For these, the locational data are easily obtained and the data are written in free format or coded onto specially prepared data sheets for subsequent transfer to cards, tape or disk.
Optical character readers may also be an aid to data input. These accept printed characters and eliminate the step of transcribing from field forms to punch cards. The disadvantages of this method are the expensive equipment required for reading, the bulkiness of the input and rejection rates due to untidy forms. A similar system is optical mark recognition. A multiple choice form is used with the selection marked.
Another technique is the use of portable data entry terminals which are battery operated, with the data recorded onto magnetic tape cassette. The data can be transferred by telephone line (acoustic couplers) to an off-line receiver which stores the data on magnetic tape or punched paper tape. This is then processed on the computer. The feasibility of these units has not been assessed in Australia. Outside field collection, they would seem to have use for field checking of data and also in the office. Photo-interpreters would code direct onto magnetic tape cassette rather than write onto date sheets. Transcribing and card punching errors would be eliminated.
Where thematic maps are available or can be prepared, a digitizer may be used to input the boundaries, whether it is to store the

information as such or to manipulate it into a form for cellular storage.

Over the last couple of years, many brands of digitizers have become available. In evaluating a digitizer configuration the following points should be considered.
- The digitizer should be interactive through a visual display unit to a mini-computer or large desk-top calculator. Instructions can be given to the operator and editing of the digitized data can be done interactively. Attribute data can be entered and verified at the same time.
- Digitizing is a slow process. Therefore the computer interfaced to the digitizer does not need to be a fast unit. As long as it has the capacity to maintain facilities for interaction and has an acceptable response time, it should suffice.
- The resolution of the digitizer should not exceed the capacity of the operator. It is ridiculous to specify a digitizer accurate to 0.02 mm when the lines that are being digitized are 0.1 mm thick.
- The digitizer should have a free moving cursor or scribe for the extraction of small irregular units and should be able to digitize in single and continuous mode (time mode being preferred to distance).
- The operator should be conversant with mapping techniques and should not be employed solely as a digitizer operator.

The hardware available for analysing satellite imagery is limited at the present time. There is an interactive GE Image 100 computer available which utilizes the digital information on the magnetic tapes directly rather than the photographic images.

There is available in Australia an image analysing computer, the Quantimet 720, which scans photographic images for tonal differences; up to 64 shades of grey may be detected. The system is interactive and can be interfaced to a mini-computer or a large desk-top calculator.

References

Beckett, P.H.T., R. Webster, G.M. McNeil and C.W. Mitchell (1972). Terrian evaluation by means of a data bank. Geog. J. 138: 430-456.

Cook, B.G. (1975). A computer data bank in a regional land use study. In URPIS-THREE: Proceedings of the Third Australian Conference on Urban and Regional Planning Information Systems, Newcastle.

Cook, B.G. and B.V. Johnson (1973). A computer data bank for regional planning. In URPIS-ONE: Proceedings of the First Australian Conference on Urban and Regional Planning Information Systems, Newcastle.

Cram, A.A. (1975). Case studies in the use of an urban and regional planning system. In URPIS-THREE: Proceedings of the Third Australian Conference on Urban and Regional Planning Information Systems, Newcastle.

Emery, K.A. (1975). Soil conservation and satellite imagery. J. Soil Cons. N.S.W. 31: 158-171.

Hoeske, F.G. and R.V. Wilson (1974). The FORINS information system for land use studies. In Proceedings of the Seventh Triennial Conference of Institute of Foresters, Australia, Caloundra, Queensland.

Humphries, B. (1973). A land information system for planners. In URPIS-ONE: Proceedings of the First Australian Conference on Urban and Regional Planning Information Systems, Newcastle.

Nichols, J.D. (1975). Characteristics of computerized soil maps. Soil Sci. Soc. Am. Proc. 39: 927-932.

Webster, R. and P.A. Burrough (1972). Computer based soil mapping of small areas from sample data. (Two parts). J. Soil Sci. 23: 210-234.

Data storage and retrieval

H.G. Mackenzie and J.L. Smith, Division of Computing Research, CSIRO, Canberra, A.C.T.

Introduction

For the soil scientist, or for that matter any scientist, dealing with an information storage and retrieval system primarily suggests the accessing of a data base for a pertinent subset of its stored information, for the purpose of a direct display or input to some other computational process. To this end the scientist's sole concern at the data base interface should be the availability of a suitable language for expressing his information requests and the mode of submitting his request and receiving the retrieved information. Should the data base be small in volume and lack complexity in the information represented in the stored data, well-established computer technology can readily satisfy these requirements, providing simultaneous access for a number of users with immediate response for most requests.

When size and complexity are both inherent in the data base it is necessary to call upon the presently evolving technology of data base management systems (DBMS). Such systems attempt to solve a number of conflicting problems which arise in large integrated data bases which are subject to continual change. It is almost certain that any presently available system has compromised on some aspects, and this will make today's implementation of a large soil information system less than ideal. This paper briefly illustrates a desirable level of data base interface for the scientist which is feasible within the decade, compares it to a typically available interface, and in the process some of the problem areas in data base management are exposed.

Data base specialists make an important distinction between information and the various levels of data. Information is our view of a part of the real world, which is formalised into collections of entities having properties which can be quantified and which relate to other entities in numerous ways. Data is the symbols (on hard copy or in computer storage) which are used to represent entities, their properties and relationships. A simple data base may represent just one collection of like entities each identified by a number of attributes, with no explicit representation of relationships except some arbitarily defined ordering in the data base. Complexity can arise when the information embraces numerous entities (e.g. survey groups, survey sites, site strata measurements, site vegetation measurements). Then if the DBMS permits, the data base designer may elect to explicity store some relation-

ships and add other storage structures which assist in versatile retrieval of subsets of the data.

The most basic relationships are classification (e.g. sites by survey group, vegetation measurement by survey group) and hierarchial classifications (e.g. strata measurement by site by survey group). The more powerful systems allow the designer to represent associations; these relate like or unlike entities in n:m mappings (e.g. geographical subdivisions to survey groups, survey sites to adjacent survey sites). These systems also support storage structures, such as indexes, which allow entity sets to be efficiently and quickly subsetted according to attributes, and other structures which allow the attributes of one particular entity to be rapidly retrieved.

There are numerous ways in which relationships can be represented in data. Increasing the size of a data base will demand that more sophistocated techniques which provide the best compromise between storage economy, speed of access, and maintenance are used. Similar remarks apply to indexing and access methods which must make the best use of present mass storage characteristics and the channels connecting them to the central computer in a general purpose system.

If a large complex soil information data base is to be successful its design, creation and maintenance must be undertaken by data base specialists. While this should not concern the scientific user, at least those scientists responsible for the project should appreciate the problems involved. What does concern all is that the DBMS chosen has to provide the tools for both the data base specialists and the range of scientific users to carry out their diverse tasks. The provision of good tools at all levels has not yet been achieved by the technology.

Experience suggest at least several different types of use of a soil information system. There are highly computational operations which use large amounts of numeric data from the data base and may simply require some suitable sequential scan of part of the data base. The same data and additional textual data will be accessed in unpredictable ways for various other purposes, some requiring direct interactive response to the scientist, others feeding the retrieved data to statistical processes, simulation processes, etc. Also there is the need to continue the input of data to the data base, either from new samples or for the correction of previous entries or the feeding back of derived results.

It follows from all of the above that the single most important feature for the scientist is a suitable language to query and update the data base. One can identify three major language levels which may be the vehicle for this function. At the highest level he could frame a query in natural language thereby invoking a number of relationships amongst real world entities in a complex

logical expression. Many such queries can be expressed just as concisely at the mathematical level of set theory and logic. This second level employs a formal model of reality which is no more than sets. At the third level of language the data model is strongly flavoured with data structures designed both for performance and to represent relationships. While there has been considerable research in the automatic recognition of natural language there has been little effort to involve the results with data base technology. On the other hand, set theoretic languages have been subject to a number of development projects and offer the best prospect in the near future for a data base interface acceptable to the scientist. Currently available languages are at the third level and, although some are of significantly higher level than others, they all exhibit much data structure dependence.

Relational data and high level data languages

A typical soil information data base may deal with a soil survey project. The entities involved are: (1) survey groups taking measurements at (2) sites selected for sampling (3) borehóle strata and (4) vegetation; as well as being located by latitude and longitude sites belong to (5) geographical units bounded by natural and man made lines and to (6) land subdivisions. At a very early stage this conceptual model would be formalised into high level data structures such as shown in figure 1 (for the time being we will concentrate the example on entities 2, 3 and 4). Here each entity is associated with a relation based on domains which are either entity identifiers (e.g. site-number) of properties corresponding to recorded measurements. In the case of the SITE relation certain properties are themselves entity identifiers (e.g. geographical unit) and thus the prospect of expansion of the model is obvious.

This relational model allows the survey information to be represented in row matrix form, any one row of a matrix being an n-tuple of attributes corresponding to a particular entity instance as shown in figure 2. A number of different query types are now illustrated using this data base and two important developmental languages having such a relational data model. (Clearly relations define sets of n-tuples). One language is called SEQUEL (Structured English Query Language) (Chamberlin and Boyce 1974) and this uses a block-structured format of English words; the other language ALPHA (Codd 1971) uses a more highly structured format and is based on relational calculus using quantifiers 'for some' ($\exists$) and 'for all ($\forall$).

Query 1

List the site-numbers of sites surveyed by the UQ1 group.
The SEQUEL expression for this query is given below and its format needs no explanation.

```
SELECT                          SITE-NUMBER
FROM                            SITE
WHERE                           SURVEY-GROUP = 'UQ1';
```

The general format of an ALPHA language statement is:
operation workspace target-list : qualifier
where the qualifier is optional. The expression for the above query is

```
GET W SITE.SITE-NUMBER    :    SITE.SURVEY-GROUP = 'UQ1'.
```

Query 2

List the strata above 1 metre at sites A2736 and A2741.
SEQUEL:

```
SELECT  *
FROM    STRATA
WHERE   SITE-NUMBER IN (A2736,A2741)
AND     DEPTH < 1;
```

ALPHA:

```
GET W STRATA: (STRATA.SITE-NUMBER='A2736'V STRATA.SITE-
NUMBER = 'A2741') ∧ (STRATA.DEPTH < 1).
```

Both languages allow complex Boolean expressions to occur in the retrieval condition. In SEQUEL the IN sub-clause allows a convenient expression of set inclusion as opposed to the conventional long hand expression in ALPHA.

Query 3

List the vegetation data recorded between latitude 27°S and latitude 28°S.
SEQUEL:

```
SELECT      *
FROM        VEGETATION
WHERE       SITE-NUMBER IN
            SELECT SITE-NUMBER
            FROM   SITE
            WHERE  LATITUDE ≥ 270000S
            AND    LATITUDE ≦ 280000S;
```

ALPHA:

```
RANGE  SITE S
RANGE  VEGETATION V
GET W V : ∃S  ((S.SITE-NUMBER = V.SITE-NUMBER)
∧(S.LATITUDE ≥ 270000S)  ∧ (S. LATITUDE ≦ 280000S).
```

```
SITE   (SITE-NUMBER, LONGITUDE, LATITUDE, ELEVATION, LAND-USE - - - -,
       GEOGRAPHICAL-UNIT, LAND-DIVISION, SURVEY-GROUP, INSTITUTION,
       LEADER, - - - - -, DATE, TIME)

STRATA  (SITE-NUMBER, DEPTH, PH, CLAY, - - - )

VEGETATION    (SITE-NUMBER, GENUS, SPECIES, AGE, CLIMATE, - - -)
```

Fig. 1 Definition of relational data model.

SITE

SITE-NUMBER	LATITUDE	SURVEY-GROUP	LEADER
A2736	273000S	UQ1	JONES
A2737	273100S	UQ1	JONES
A2738	273200S	UQ1	JONES
A2741	283100S	CS1	SMITH
A2742	283200S	CS1	SMITH

STRATA

SITE-NUMBER	DEPTH	PH
A2742	0	6.3
A2742	.1	6.3
A2742	.2	6.2
A2742	.7	6.7
A2741	0	6.2

VEGETATION

SITE-NUMBER	GENUS	SPECIES	AGE
A2742	Limnodynestes		
A2742	Litoria		
A2742	Uperoleia		

Fig. 2 Relational data base

In the SEQUEL expression the set inclusion construct is used in a nested expression causing the output of one selection mapping to become the input of another.

In ALPHA the expression reads more as an iterating cross-referencing operation between two relations. This requires the definition of tuple variables (S and V) by RANGE statements which specify the range over which they are to iterate. The quantified S reads for some S (or at least one S) within the iteration range. The order of occurrence of tuple variables defines the nesting.

Query 4

List all vegetation by botanical name and evaluation at which they were sampled.

SEQUEL:

```
SELECT      GENUS, SPECIES, Q
FROM        V IN VEGETATION
COMPUTE     Q =
SELECT      ELEVATION
FROM        SITE
WHERE       SITE-NUMBER = V.SITE-NUMBER;;
```

ALPHA:

```
RANGE  SITE S
RANGE  VEGETATION V
GET W  V. GENUS, V.SPECIES, S.ELEVATION
       : V. SITE-NUMBER = S.SITE-NUMBER.
```

In the SEQUEL expression V is known as a correlation variable and is similar in purpose to the ALPHA tuple variable; Q is known as a computed variable and it is necessary to use this when all the data desired cannot be selected from one relation. In these cases the ALPHA expression is considerably simpler, once the range statement is understood.
It should be clear from these examples that a data base could be created and updated using language expressions of similar structure (for examples see Codd (1971)).

Transformation of the relational model

Using the concepts of join and projection of relations (Codd 1972), one collection of relations can be transformed into an arbitrary number of meaningful relations. For example the relation VEGLOC (genus, species, latitude, longitude, elevation, climate) is easily derived from two relations in figure 1. Should this external model of the data base of figure 2 be provided for the user, the expression for Query 4 in Section 2 would become considerably simpler.

Different external models of a data base have been provided by DBMS for some time (see sub-schemas in CODASYL (1971) and program specification blocks in IBM (1974)), but only for low level data models. At the relational level the concept means that, even if the data was stored in the data base according to the internal model of figure 2, the user could program certain operations (e.g. queries) as if the relevant data were stored in the row matrix defined by VEGLOC above. The DBMS is then responsible for mapping the particular external model and translating the accompanying query into the equivalent data base operation.

Additional restrictions must be imposed on external models which are used for creation or update. In any data base relation the concept of entity identifier (called primary key in Codd (1972)) must exist, this providing unique identification for any row - n-tuple. For example in the VEGETATION relation of figure 1 this

identifier would be the combination of SITE-NUMBER, GENUS and SPECIES, neither property being sufficient by itself. Any new data added to a data base must contain these tuple identifiers and the DBMS would reject any entry with a duplicate identifier. The relation VEGLOC offers several difficulties if used as a model for update. It contains data belonging to two different data base relations but does not contain the data necessary to identify either nltuple or make otherwise incomplete entries.

The problem of missing data can be overcome (by system or user) but the lack of identifiers cannot, and data from the VEGLOC relation cannot be presented for storing in the data base. This simple example illustrates one feature which makes data base creation and update much more onerous than query. The user must be provided with an appropriate external model, or sufficiently understand the entities and entity relationships involved in the data base so that he can present meaningful data.

Normalisation

Because of certain data dependencies which exist, the relations of figure 1 are not an ideal internal model of the data base. The process to be exemplified here of transforming a conceptual relational model into a suitable internal model is called normalisation (Codd 1972), and would be carried out by the data base designer. One reason for applying this process is that certain undesirable data instance dependencies can be presented. Also it leads to the elimination of much data redundancy in a large data base and thereby to improvements in performance.

In the SITE relation of figure 1 SITE-NUMBER is the unique identifier and the properties SURVEY-GROUP, INSTITUTION and TEAM-LEADER have the intended purpose of describing the one survey team which carried out the measurements at a site. Thus for each SITE-NUMBER there is one and only one SURVEY-GROUP and for each SURVEY-GROUP there is one and only one GROUP-LEADER and INSTITUTION. However one SURVEY-GROUP is associated with more than one SITE-NUMBER. This transitive dependence amongst properties in a relation always indicates that the relation should be replaced by two projected relations in the data base model. These relations and the data equivalents of figure 2 are shown in figure 3.

Figure 3 illustrates that in a large data base a large number of redundant entries for LEADER, INSTITUTION and other survey group properties are avoided. Furthermore the dependence of data describing survey groups on the existence of some of their survey data is removed. With the new relations once the entity identifier SURVEY-GROUP is known the associated properties describing the survey group can be entered in the data base, and the survey results entered by a completely independent job or person.

The scope of the model in figure 1 is more limited than the soil information data base described at the beginning of section 2. The

SITE (SITE-NUMBER, LONGITUDE, LATITUDE, ELEVATION, LAND-USE, - - -,
GEOGRAPHICAL-UNIT, LAND-DIVISION, DATE, TIME)

SURVEY (SURVEY-GROUP, LEADER, INSTITUTION, - - -)

SITE

SITE-NUMBER	- -	LATITUDE	- -	SURVEY-GROUP	DATE	TIME
A2736	- -	273000S	- -	UQ1	- -	- -
A2737	- -	273100S	- -	UQ1	- -	- -
A2738	- -	273200S	- -	UQ1	- -	- -
A2741	- -	283100S	- -	CS1	- -	- -
A2742	- -	283200S	- -	CS1	- -	- -

SURVEY

SURVEY-GROUP	LEADER	INSTITUTION	-
UQ1	JONES	-	-
CS1	SMITH	-	-

Fig. 3 Normalised relations

only data reference to geographical unit and land division entities is by the inclusion of their entity identifiers as a property in the relation SITE. If additional properties of these entities were included in the relation SITE, there would be need for further normalisation

Network data structures

With presently available data base technology it is unlikely that the data base designer could consider his design complete after normalising the relational view. This is because much of the data base will involve cross-referencing between relations, for example Queries 3 and 4 in the second section (p. .). In Query 3 the cross reference is made via an equality test on SITE-NUMBER having selected a SITE row tuple satisfying in the latitude constraints. To perform this operation on the matrix data structure of figure 2 requires a search of both arrays with cost proportional to the number of entries. When the dimension of these arrays is very large the search times cannot be tolerated for the frequently repeated types of operation, and the data base designer must add additional structure to his data base to reduce the time and cost involved.

The CODASYL specifications (CODASYL 1971) typify the data structure techniques available in present DBMS. A record type is defined in correspondence to each normalised relation. A record instance is then stored in the data base in correspondence to each row n-tuple of a relation. In addition a number of relationships

are defined to correspond to the type of cross referencing described above. In CODASYL terminology this structure is called a set but to distinguish it from the very different mathematical logic concept it is sometimes called a coset (Nijssen 1974). A coset structure consists in part of a 1:n mapping between one record instance (the coset owner) and n record instances (the coset members). The owner and member record types must be different. The data language allows the user to retrieve the owner record and member records of a coset instance in a design specified order. The DBMS structures a data base (for example by using pointers) so that once any member or the owner of a coset instance is located all the other members can be retrieved in order with minimal time and cost. The structure resultant from numerous cosets is literally a network data structure.

If we consider a data base designed at this level corresponding to the relations of figure 1, normalised, ther would be four different record types represented by the boxes in the data structure diagram of figure 4. Each record has an internal data structure consisting of data items corresponding to the property domains of the relation. The records are related by three cosets, represented by the directed lines joining the boxes. The first coset, called SITE-SET, is used to relate a SURVEY record instance to the SITE record instances for any particular survey group. Thus the data structure embraces all cross references in either direction at the relational level between the SITE relation and the SURVEY relation which use an equality condition on SURVEY-GROUP (see figure 3). It follows that it is unnecessary to store a data item for the property SURVEY-GROUP in any SITE record. The other two cosets are defined similarly using equality on SITE-NUMBER and the total structure is the familiar hierarchial classification of figure 4.

The same rationale can lead to network structures in a data base embracing more entity types. An example is shown in figure 5 where several hierarchies and associations between a number of entities are represented. Not all relationships represented by cosets are based on identities, for example the relationship between GEOGRAPHIC GRID and SITE would be computed by an application program.

Available data languages in data base management systems

Languages which have a relational data structure have only been researched in the last five years and are not implemented in current DBMS. There are several reasons for this, one being that such languages are inherently very powerful and allow queries which are very demanding in data base processing to be expressed with ease. In addition the problem of translating statements in these languages to existing data manipulation languages which have a network data structure is made difficult by data structure dependencies which exist at the latter level. Such data dependencies also affect those higher level query/update languages which are available (see next section, p30).

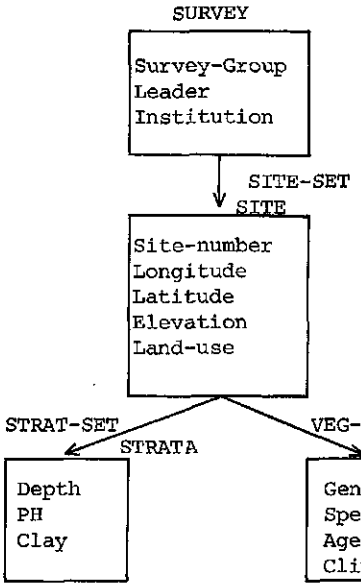


Fig. 4 Hierarchical data structure diagram

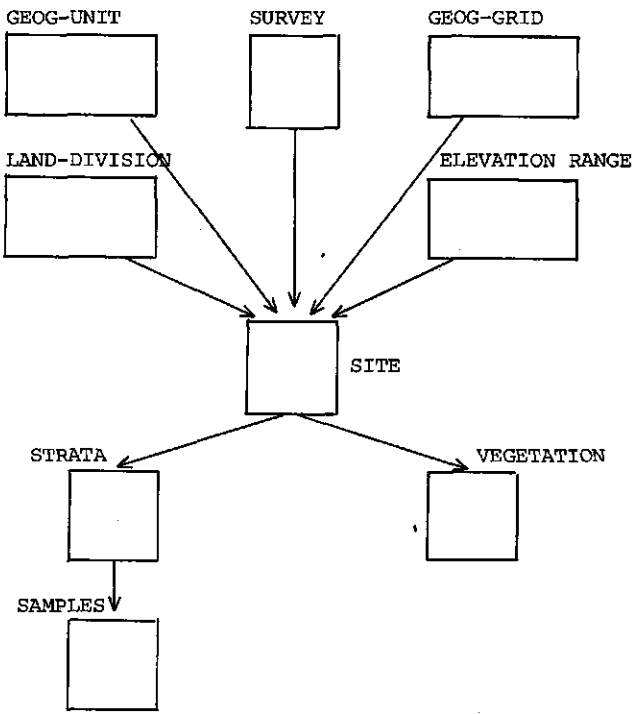


Fig. 5 Network data structure diagram

All DBMS provide a language equivalent to the CODASYL Data Manipulation Language (DML) (CODASYL 1971). This is a general purpose data language in which all data base operations on hierarchical or network data structures can be programmed. This language is the likely choice for programming large scale data base operations such as creation. The commands allow individual records to be stored or retrieved, and certain qualifications can be expressed with a record retrieval command. When using any command to store a record, account must be taken of all the coset structures involving the record and this task can be complicated in a network data base. For example when storing a SITE record in a data base described by figure 5, there are five coset relationships which must be specified or deferred for later specification. If deferred, the record cannot be retrieved using that relationship until it is specified.

A DML is always provided in the framework of a host language (typically COBOL, PL/1 or FORTRAN) so that retrieved data can be immediately processed or buffered using the facilities of the host language. This may be an important requirement and it should be noted that in some DBMS the higher level languages cannot be used in the host language framework.

Because of numerous complexities associated with the general network data structure, many DBMS support only a limited class of structures. The most common limitation is to restrict the total data base structure to be hierarchial (e.g. figure 4). SYSTEM 2000 (CDC 1974), which is available on many computers, has this restriction. TOTAL (CDC 1974), which is also available on many different computers supports a limited network structure in which a record type may be an owner of cosets or a member of cosets, but not both. In IMS (IBM 1974) a more general network structure is supported but the DML is restricted to processing derived hierarchical structures, and the commands are transformed by the system into network operations. In FORDATA (Smith and MacKenzie 1974) an unrestricted network structure is supported. The restriction to hierarchical structures does make a DML considerably simpler, but the commands are always limited to manipulating one tree instance at a time.

The higher level query/update languages have a syntax much more like the relational data languages of the second section (p. 21) but they differ significantly in that they operate on the hierarchical (or restricted network) data structures. Thus the user must be aware of the coset structure in the particular data base design. They are more powerful than the DML in that a retrieval request can produce an arbitrary number of hits and involve the processing of a large number of records. Retrieval conditions can be complex Boolean expressions but the quantifiers of the predicate calculus are not supported. In those systems where the data base is restricted to a hierarchical structure some difficulties in these languages are avoided because of the relative simplicity of the data base. However the expression for a given query can differ markedly depending on whether a cross reference is supported by a coset structure or not. An example is given in the next section.

In this context an important property of DBMS concerns the facilities for designing index structures into the data base. Many systems do not support indexing at all, so that many retrievals will involve exhaustic searches of the data base. Thus while the expression of the retrieval request may not differ significently between two systems the time and cost of execution will. For example the CODASYL specifications (CODASYL 1971) define indexing facilities and these are provided in the FORDATA implementation (Smith and MacKenzie 1974), but TOTAL (CDC 1974) does not.

Query/update language

An example of a high level query/update language which can be used with TOTAL data bases is ATHENA (CDC 1975). In figure 6 three TOTAL data structure designs are proposed for a similar data base example to that discussed above. Note that because of the restrictions on the use of cosets, the data structures of figures 4 and 5 are not permissable in TOTAL. A statement in ATHENA for Query 3 of the second section (p. 22) and the data structure of figure 6 (i) is:

```
FORMAT, SITE-NUMBER OF SITE, GENUS, SPECIES, AGE, CLIMATE, WHERE
LATITUDE > 270000S, AND LATITUDE < 280000S
```

[1]

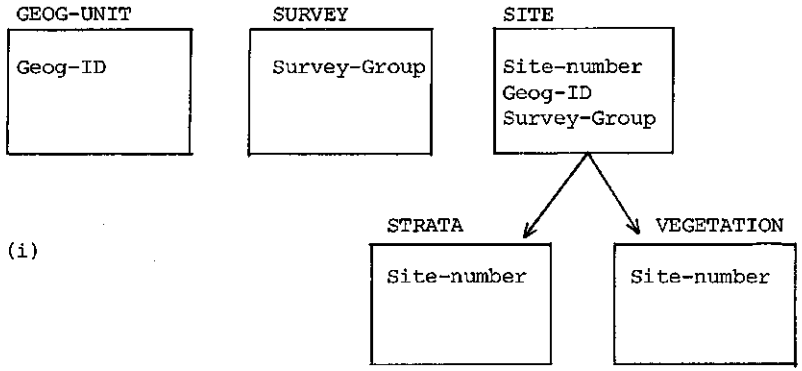
The expression for $\leq$ is more verbose and is not worth elaborating on here. The format of ATHENA statements of this type is actually simpler then the equivalent expressions in SEQUEL and ALPHA (see the second section, p. 22). The reason lies in the coset structure dependence of ATHENA statements. Cosets can only be defined on equality of a data item in the member record with the primary key of the owner record, and this meaning of the coset structure is involved implicitly in ATHENA by including data items from SITE and VEGETATION records in the above query. On the other hand in relational languages the cross reference is handled in the language syntax, and hence the statement is more complex.

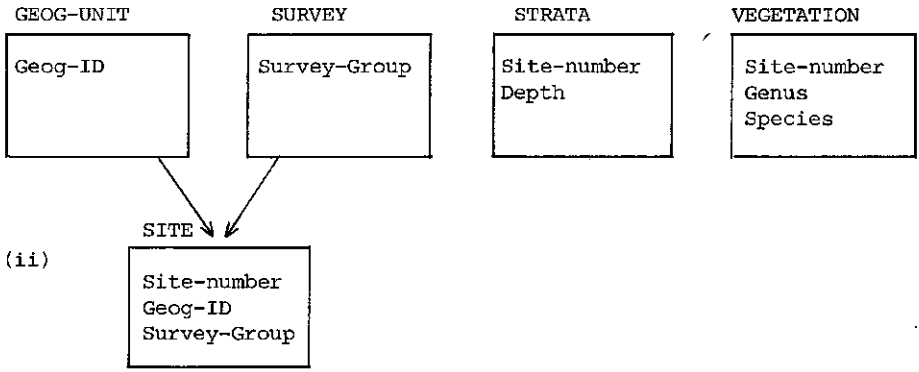
Should the data base design conform to figure 6 (ii), so that there is no coset structure which can be invoked for the purpose of this query, the essence of its formulation is given below.

```
SUBSET, *FORMAT, SITE-NUMBER OF VEGETATION, GENUS, SPECIES, AGE,
CLIMATE, WHERE, SITE-NUMBER OF VEGETATION =*, SITE-NUMBER OF SITE,
*.*, WHERE > 270000S, AND LATITUDE < 280000S.
FINISH.
REWIND, SUBSET
PROCESS 1.
```

The result of the above SUBSET command is to build a file of records that each contain a command of the form:

[1] Certain ATHENA/TOTAL rules concerning names of data items have not been observed here.





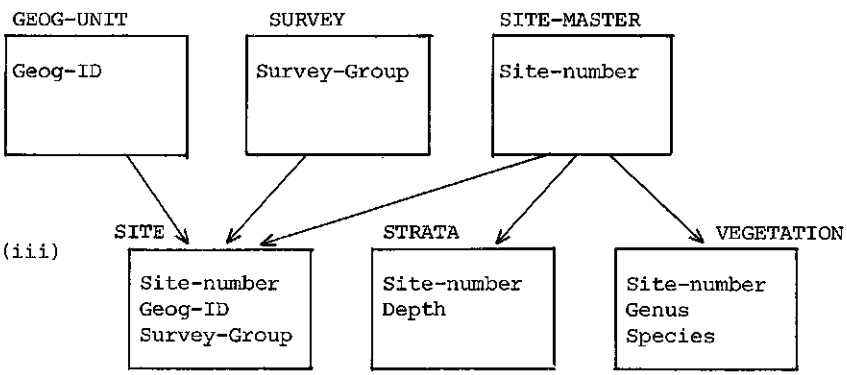


Fig. 6 TOTAL data structure designs

FORMAT, SITE-NUMBER OF VEGETATION, GENUS, SPECIES, AGE, CLIMATE WHERE, SITE-NUMBER OF VEGETATION = XXXXXX.

There are as many records in the file as there are SITE records satisfying the latitude criterion and the particular site number occurs in the place of XXXXXX. The PROCESS command causes these commands to be submitted in order, thus producing the desired results from the data base.

The data structure of figure 6 (iii) contains the coset structure of figure 6 (ii) as well as additional coset structures and a bridging record called SITE-MASTER. The latter allows the above query to be expressed more simply and executed more efficiently than with the limited structure of figure 6 (ii). The expression is given below.

FORMAT, SITE-NUMBER OF SITE, GENUS, SPECIES, AGE, CLIMATE, WHERE, LATITUDE > 270000S, AND, LATITUDE < 280000S, AND, SITE-NUMBER OF SITE-MASTER.

The INSERT command allows a new record to be entered into the data base with a complete or partial list of data items stored. If the record is a coset owner its primary key data item must be provided; if the record is a member of one or more cosets, the primary keys of all its cosets owners must be provided through the designated data items in the member record, as illustrated below for figure 6 (iii).

INSERT, SITE-NUMBER = A2739, GEOG-ID = SW12, SURVEY-GROUP=UQ1, WHERE, LINK TO GEOG-UNIT.

Note how the coset restrictions which limit the type of network structure keep the form of these statements reasonably simple. Similarly CHANGE and DELETE commands allow data items to be modified and records to be deleted.

ATHENA also provides a number of commands for interfacing to the data base, output control and arithmetic functions such as AMEAN, STDEV, COVAR. TOTAL does not provide any indexing structures and so the performance of any command not based on a primary key will deteriorate rapidly with data base size.

Network data base design and maintenance

The examples of the previous sections have illustrated the main data structuring technique for data base design. However network data base design and maintenance involves numerous other important tasks which must be attended to by the data base specialist. These tasks are made difficult with size, complexity, unpredictable use and frequency of update of the data base. The better the system and the data base design the less effect these problems will have on the scientific user. However with present technology they are certain to be felt in some ways, as has already been illustrated

with data structure limitations and data structure dependence. Poor design and maintenance of a data base will soon make itself felt in time and cost to the user. Some of the problems are briefly indicated here.

It is necessary to subdivide large complex data bases into areas (CODASYL 1971)-hence the term multifile system. An area usually contains records of the same type or of a small number of different types; in the latter case the records would probably be related by cosets such as SITE, STRATA and VEGETATION in figure 6 (i). The area internal organisation must adapt the characteristics of disc storage devices and huffering to high speed memory, to the predicted use pattern of the records. Hence each record is prescribed a primary access mode called its location mode (CODASYL 1971). For example the primary key, SITE-NUMBER for SITE records, would be the likely basis for most record retrievals, and so the location mode for SITE records would be designed on SITE-NUMBER. The most common methods of transforming primary keys to area addresses are by use of indexed sequential organisation or by calculated addressing. Each method provides certain advantages, but a primary requirement is to maintain efficient access after a reasonable amount of modification, deletion and insertion of records to the area.

If an indexed sequential organisation is used the bulk of the area must be created by presenting records for storing in sorted order or primary key, and records are then placed sequentially in area addressing space. A very efficient index can then be constructed by the system to support random retrievals, but at the same time the complete set of records can be processed efficiently in key sequential order. Calculated access relies on a transformation algorithm which can map key values to the area addressing space with a uniform density. In using this method the designer trades the advantages of sequential placement for better random retrieval performance which is durable under much updating.

The location mode design becomes more complicated when coset structures have to be involved. For example the most common use of STRATA records might be to access them in relation to SITE records. Then the most desirable placement for STRATA records in area addressing space can easily conflict wit the placement of the totality of SITE records. Clearly with a data base such as that represented by figure 5 many compromises must be made in choosing the area/location mode design.

It is also necessary in data base design to take advantage of logical orderings of records within the various classifications which are represented in the data structure. For example, within a coset instance an ordering can be defined which is temporal (i.e. first in first out or last in first out) or sorted by the values of defined data items. The DML allows the user to retrieve the records one at a time from the coset in the designed order. Application programs written in DML can then take advantage of

(and become dependent on) this order. For example applications which process STRATA records according to DEPTH can take advantage of STRAT-SET in figure 4 being ordered on ascending values of DEPTH.

It is likely that certain data items which are not associated with the location mode will also be identified as the basis of frequent retrieval requests. For example, with any of the above data base examples VEGETATION may be frequently retrieved on the basis of GENUS and SPECIES. The designer must decide whether the request frequency warrants the construction and maintenance cost of an index to speed retrieval requests. There are numerous ways of constructing indexes some of which can be associated with coset order (see SORT KEYS and SEARCH keys in CODASYL (1971)), but the accumulation of too many indexes in a rapidly changing data base cannot be afforded.

The task of maintaining a data base so that it is available to the scientific user whenever required should not be underestimated. It should be apparent that the stored network structure is extremely complicated and should it be corrupted it would be difficult to correct it using normal data base languages. Because corruption can occur through no fault of the data base manager or the user, as well as through their own error, it must always be anticipated and a recovery method prepared. Clearly recreation of a large complex structure would be overly expensive in both cost and delays incurred, even if the original input data were all available. Backup copies of the entire data base must therefore be maintained and if update activity is confined to certain areas these may be copied more frequently. However with frequent update it is not feasible to maintain up-to-date copies and some form of journal must be maintained which, when used in conjunction with a recent area copy, allows the area to be restored. Different types of journal may be kept to combat different error situations. For example it may be that a corrupted data base can be restored to a recent non-corrupt state by over-writing known updated portions from a journal of 'before-update-images'. In these situations any user involved in updating the data base may be forced to restore data. The task of ensuring the integrity of a data base demands proper management procedures for updating, and the saving of journals and area dumps.

Restructuring and growth of a data base are two difficult problems of data base management. Continued updating conforming to an existing design will often lead to deterioration in performance and scheduled reorganisation within the same design will be required. Changing patterns of use may require alteration to the design. Some design alterations may not affect users (e.g. the creation and deletion of indexing structures) except in cost and response time. Other design alterations which are data structure changes or changes in logical ordering may require the reprogramming of many applications as well as the reorganisation of the data base. In extreme cases the redesign may correspond to the growth of the

information scope (e.g. figure 4 to figure 5) or equivalently the integration of separate data bases. Data structure dependencies are likely to affect even very high level language programs in these cases.

Conclusion

Ideally the scientific user of a data base should be provided with a convenient interface which allows him to store and retrieve information with little time and effort. Research and development presently under way is aiming to provide this interface using so called relational data bases. At this level the user is provided with a language which is structured English or more obviously based on mathematical logic, and a data model which is no more than sets of n-tuples. The actual stored structure of the data base is of no concern to the user, although it will be changed independently of the user model in order to meet changing performance requirements.

Currently DBMS store a network structured data base and attempt to provide a high level user interface through so called query/update languages. Query conditions in these languages are limited to Boolean expressions and the data model is a hierarchial or network record structure. Most systems are limited in the complexity of the data model they can store (e.g. only hierarchial) and various data structure dependencies force the high level language user to be aware of the exact network structure being used to represent the information.

Performance requirements in time and cost make the design of a large complex data base a difficult problem and there is need for continued maintenance and redesign. In order to achieve maximum efficiency certain data storage and retrieval operations have to be programmed in lower level data manipulation languages, in which advantage can be taken of the detailed design. The need for both high level and low level language likely to always exist.

A number of other requirements can also be identified concerning scientists' data storage and retrieval. Some of these are outside the presently accepted scope of DBMS and should be found in the network operating system. The latter should allow the scientist to use the information system in batch or interactive mode with multiple access by a number of scientists. Suitable terminals are required for display of processed data in the form of graphs, histograms, etc. The DBMS should provide the host language environment, and interfaces to report generators and other large systems such as simulation systems.

References

Chamberlin, D.D. and Boyce, R.F. (1974). SEQUEL : A Structured English Query Language. Proc. of 1974 ACM SIGFIDET Workshop, Ann Arbor, Michigan, April 1974.

Codd, E.F. (1971). A Data Base Sublanguage founded on the Relational Calculus. Proc. 1971 ACM SIGFIDET Workshop, San Diego, Calif., Nov. 1971.

Codd, E.F. (1972). Further Normalisation of the Data Base Relational Model. Courant Computer Science Symposium 6, in Data Base Systems (ed. by Rustin), Prentice-Hall, 1972, pp. 33-64.

COSDASYL (1971). Data Base Task Group Report, ACM, New York, 1971.

IBM (1974). Information Management System / Virtual Storage (IMS/VS), Application Programming Reference Manual 5740-XX2.

Nijssen, G.M. (1974). Data structuring in the DDL and Relational model. Data Base Management, Cargese (Corsica) April 1974. (ed. by Klimbie and Kofferman) North Holland, 1974.

CDC (1974). SYSTEM 2000 Level 2 User Information Manual, 76074000.

CDC (1974). TOTAL-CDC Reference Manual Version 760300.

Smith, J.L. and H.G. Mackenzie (1974). FORDATA: A data base management package under FORTRAN on Cyber Computers. CSIRO, Division of Computing Research P.O. Box 1800, A.C.T.

CDC (1975). ATHENA, On-Line Interactive Retrieval/Update Language, 76071400C.

Land resource information systems: use and display

B.G. Cook, Division of Land Use Research, CSIRO, Canberra, A.C.T.

Introduction

The term *information system* implies an organized collection of data, with procedures for its use which go beyond mere re-presentation. The term implies a purpose : data is chosen and organized, and the use procedures developed, towards that purpose.

Much land resource inventory is undertaken without a specific purpose in mind, perhaps with a bias towards one or several application fields, but with the aim that some generality of usefulness will be achieved. The general purpose inventory is thus at odds with the requirements of an information system: without an underlying purpose there is no basis for the data organization or use procedures of the information system.

The success of an information system, as measured against its purpose, must depend on three main factors
1. the appropriateness and quality of the data,
2. the organization of the data,
3. the use procedures.
This paper seeks to examine and comment on some problems and techniques in using and displaying computer-stored land resource data. However, as data use is so dependent on data quality and data organization, some reference to these areas will be unavoidable.

Much of what follows is as relevant to a manual system of information storage as to a computer-based information system. The advantages of the computer system are not so much in providing a storage medium as in allowing efficiency and versatility in application and display.

Modes of use

We may distinguish between a number of modes of use of a landresource information system. The most straightforward is the selective retrieval and presentation of data stored: an apparently trivial procedure but nevertheless one which is a considerable advance on non-computer systems of storage, for which data access and display are laborious processes. A significant advance is achieved by the ability to compute from the raw data stored some function of the data values, with the intention of deducing properties or capabilities of the land not explicity

stored as data. Thus, in the South Coast Project of the CSIRO Division of Land Use Research (Cook 1975), algorithms predicting suitability of land for various uses (agriculture, forestry, urban use, recreation, conservation) were constructed, to operate on basic physical, biological and social data held in a data bank. For example, an algorithm designed to classify land as either arable, suitable for improved pasture or rough grazing, or unsuitable for agriculture, considered the present use of the land, its slopes, and soil properties such as texture, drainage and salinity. This approach allows the prediction algorithms to be modified as conditions and ideas change, while the basic data on which they operate remain (reasonably) constant. A similar mode of data use is the search for land satisfying certain criteria. The criteria may be simple limits on data items stored, or more or less complex functions of the data.

There may be limits, however, arising from the complexity of land and the imperfection of its description, to what can be achieved using suitability rules or search criteria. Attempts to refine the criteria used may suffer from the same difficulty as is found in another discipline concerned wit imperfectly described complex objects, document retrieval - the compromise between *precision* and *recall*. Consider a function designed as the criterion for indentifying land suitable for a particular use. The more that function is refined to exclude unsuitable land (i.e. to improve precision), the more likely it is that some suitable land will be excluded (i.e. recall will suffer); conversely, an attempt to include all suitable land (i.e. to improve recall) is likely to also include more land that is unsuitable (i.e. precision may decrease). Retrieval rules may therefore not be amenable to continual refinement, and the results of their application should be viewed in the light of this precision-recall compromise.

For some purpose the use of a 'natural' classification, determined by clustering in an attribute-space, may avoid the difficulties of precise class definition. This is especially applicable where interest is in identifying *similar* areas of land. However, land capabilities will not necessarily correlate closely with such classes: land within one class might have a wide spread about the critical value of an attribute important for particular use.

Computing with land resource data

The design of retrieval rules is made difficult by the fact that the variability of land results in the common use of complex land descriptions.

To enable land resources to be described, the land surface is delimited into units, either on a regular geometric basis, or by irregular boundaries subjectively defined in an attempt to reduce the complexity necessary in the land description. In practice, the units adopted are rarely internally homogeneous, and hence do not allow a simple description; description is usually in terms of a

pattern of simply described (but unmapped) components. (Indeed, if units were made fine enough to allow simple description, the pattern description would no longer be explicit and important pattern information would be difficult and expensive to recover.)

Attributes of land, such as landform, soil, vegetation are therefore usually described within each unit as a pattern, by describing both the pattern elements present and their interrelationships. The usual situation, then, is that much basic descriptive data stored about the delimited units of land is already quite complex in structure. This makes the design of retrieval rules a logically more difficult task than would be the case with simple descriptors.

When dealing with delimited units which are internally homogeneous in the described attributes there is no difficulty in deducing the co-occurence of, for example, a soil type with a landform type by simple logical intersection or overlay. Where attributes are described in terms of pattern with soil and landform types as unmapped elements within their respective patterns, co-occurence cannot be deduced from independent descriptions - it must be explicicty described. The same problem arises when joint consideration of independently-mapped attributes is required - where pattern description is used, only an obscure probabalistic description of co-occurence can result (McAlpine and Cook 1971).

Another important consideration is the level of generalisation of a description. Attributes of land can be described at different levels of detail, and pattern can often be seen and described at a number of scales. There needs to be a clear understanding on the part of the data user of the 'scale' at which the description is applicable. For example, the detail necessary for farm subdivision planning is unlikely to be found in data collected from a regional planning viewpoint. Conversely, detailed data is likely to be found unsuitable for regional planning without generalisation, which may not be an easy task. An information system to serve several levels of application really needs data stored at more than one level of generalization.

Efficiency

The efficiency of use of an information system may be considered from the points of view of cost, convenience and correctness of output.

Efficiency in cost terms will depend upon the medium used for data storage, the structure of the data stored, and the computional and data access efficiency of the retrieval algorithms. Convenience will depend on both the accessibility of the data stored and on the ease with which retrieval algorithms may be constructed. Ability to provide correct and useful output from the information system will depend on the quality and appropriateness of the data stored and on the appropriateness and correctness of the algorithm constituting the retrieval criterion.

The efficiency of a retrieval algroithm addressing a data base of moderate to large size will often depend more on data access considerations than on computational method. If the data base is a sequential file the aim should be to minimise the number of file passes; in a structured data base (Mackenzie and Smith 1976) the number of record accesses should be minimised. This can sometimes be assisted by preprocessing a section of the algorithm which would otherwise require repeated access to and computation with some section of the data, and by holding the results in some more accessible temporay location.

Ensuring the correctness of a retrieval algorithm is a significant problem in any application environment which requires novel algorithms for many individual retrievals. The effort required to adequately check a retrieval program which is to be used only once may be considered quite unacceptable. Probably the only solution to this difficulty is by ensuring that retrievals can be specified in as simple a manner as possible - the simpler that task the less the probability of error.

Simplicity of retrieval specification may be approached at two levels:
1. by use of a generalized data base management system for data storage; this should free the programmer from some routine data handling worries which otherwise complicate his task; and
2. by development of general retrieval programs which allow retrieval algorithms of commonly occuring forms to be specified by parameters or high level language.

Display

Communication of the result of a retrieval to the data user or to a wide audience should be considered as important as the retrieval itself. Information may be communicated in many ways - here I will confine discussion to the usual vehicle for displaying land information: the thematic map. Two forms of display will be discussed: the character matrix map produced by line printer and usually associated with a grid cell descriptive scheme, and the more conventional region boundary map produced by automatic plotter.

Usually, the aim of a thematic map is not to indicate the value of a land attribute at every point (although it can be used for this purpose), but rather to indicate the distribution of attribute values in an area; the interest is morphological rather than point-definitive. It is important then that the form of display convey a clear impression of the distribution - that areas which are different (in terms of the theme being displayed) be clearly distinguishable from one another and that connected areas of similar land be clearly seen to be connected.

The line printer character map is far from ideal as a display medium. The contrast between different characters is low and it requires some visual effort to distinguish the extent of a uniformly

mapped area. It can be improved by selective overprinting to produce a larger contrast range, and by suitable use of blanks either along or within boundaries. Boundary lines may alternatively be drawn in manually after printing. In my view, the character map is poor cartography justified by its inexpensiveness and its ease of production using straightforward computer techniques.

The more conventional style of thematic map, usually produced by automatic plotter, has the potential to be a more acceptable display medium, but a satisfactory visual effect is not achieved without care. It is desirable that connected regions of similar thematic class be merged, by not plotting boundary lines redundant to the theme. The placing of a simple label within each region defines thematic classes, but is unlikely to give a satisfactory visual effect. This can be achieved by automatically hatching each region, using a selection of suitable hatching patterns, but the method is costly and uses a great deal of plotter time. For many purposes, hand colouring of a labelled map is a satisfactory alternative. Other devices, such as the visual display screen or computer output microfilm, within the limitations of their restricted format and/or lack of permanence, can be used to advantage where available.

In those applications for which the digitising and storage of region boundaries cannot be justified, there is an alternative which appears to have been little used. Rather than plot a complete thematic map, the system produces a label overlay to a standard, conventionally produced, boundary map. The information system need only hold the co-ordinates of a suitable location for label placement in each region. Display is of course restricted to the scale of the standard boundary maps prepared, and removal of redundant lines is not possible.

The display of retrieval information by thematic map often requires that a classification be imposed upon the information to define a map legend; this should have as few classes as are needed for easy comprehension. Care needs to be taken both in selecting the class intervals and in designing the legend; either a poorly chosen classification or a carelessly worded legend can convey a false or misleading impression.

Much of the utility of a map is lost if it cannot be related to a suitabel base, yet it is not uncommon to see maps output by computer for which no corresponding base map exists. Most map series in Australia, at scales from 1:250 000 to 1:25 000, use a Transverse Mercator projection with either the old Austalian National Grid (yards) or the now-standard Australian Map Grid (metres). It is therefore important that land resource data retrieved from an information system be able to be displayed in this projection with grid and scale corresponding to an existing base map.

A particular problem arises with grid cell data mapped by line printer. The map scales which can be produced by this method are

severely constrained by the fixed print-position geometry of the printer and, unless the grid chosen conforms to the usually non-square line-column spacing ration, northing and easting scales will differ. For such a line printer map to overlay a standard base map, the grid chosen needs to be congruent with the map grid, and must have dimensions which result in a line-printer map at the required scale. Grid cells based on latitude and longitude can never be mapped by these means to overlay a Transverse Mercator projection (except at scales larger than about 1:25 000, when the geographic grid becomes sensibly linear within a conventional sheet size).

Where a plotter is used for map production, the constraints of the line printer disappear, but there remains the need to produce a map at a suitable projection. No problem arises if location is represented by map grid co-ordinates, but this will often not be the case. (The zonal discontinuities of the Transverse Mercator system, every 6 degrees of longitude in the case of the Australian Map Grid, argue against the adoption of grid co-ordinates except where interest is confined to a region within one zone).

Computer programs for transformation between geographic co-ordinates and grid co-ordinates exist, and may be used to convert information system co-ordinates when necessary to the appropriate grid for plotting. The computation is lengthy, however, and its application to each individual point location required may prove costly. An alternative is to use approximations sufficiently accurate for the map being produced. For example, a single projective transformation (Ahuja and Coons 1968) may be used to transform geographic to grid co-oridnates on a 1:50 000 scale map sheet (¼ degree square) with maximum error of about one second of arc in latitude, much smaller in longitude. If a larger area is involved or higher accuracy required, the area may be divided into a number of rectangles, each with an appropriate transformation.

Maps which display an area which crosses a zone boundary may be best presented by butting the two Transverse Mercator zones along their join line (figure 1), which requires a simple translation and rotation transformation. (This produces a slope discontinuity across the join, but allows base maps to be similarly butted for comparison.)

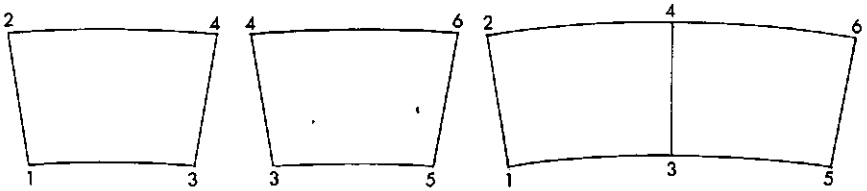


Fig. 1

Conclusion

Effective data use is both the aim and justification of a land

resource information system. The cost of designing, establishing and maintaining an information system will normally only be warranted by an expectation of sustained interest in and continuing use of the data stored. In other situations, a relatively inefficient but simple data organization will often be preferable.

Improvement in information systems will not occur automatically. Inadequates and inefficiencies, often obvious to the system user, will go uncorrected unless communicated to those responsible for data collection and system design. It is important that there exist clearly defined procedures to encourage such communication.

References

Ahuja, D.V. and S.A. Coons (1968). Geometry for construction and display, IBM Syst. J. 3 & 4: 188-205.

Cook, B.G. (1975). A computer data bank in a regional land use study. In URPIS-THREE: Proceedings of the Third Australian Conference on Urban and Regional Planning Information Systems, Newcastle.

McAlpine, J.R. and B.G. Cook (1971). Data reliability from map overlay. Paper presented at 43rd ANZAAS Congress, Brisbane.

Mackenzie, H.G. and J.L. Smith (1976). Data storage and retrieval. In these Proceedings, p.

Computer assistance in the preparation of a detailed soil survey of the Padthaway irrigation area

D.W. Armstrong, Department of Agriculture, Struan, South Australia
K.G. Wetherby, Department of Agriculture, Cleve, South Australia

Abstract

A computer has been used in the collation and analysis of field data for construction of a detailed soil map of about 12 000 ha. Field data were recorded on specially designed forms to allow direct transfer to punch cards. Obvious errors and omissions in the data were detected by validation programs and corrected where possible. Additional programs were written to retrieve information from the profile descriptions, to classify profiles using specific factors, and to map the distribution of profile types.

Background

The Padthaway Irrigation Area covers 12 000 ha in the South East of South Australia. Until the late 1960's irrigation was applied mainly to lucerne and perennial pastures, but irrigated vines and vegetables are becoming increasingly important crops.

A soil survey was requested by the local Soils Extension Officer of the Department of Agriculture to enable him to give more detailed advice about the soils in the area, particularly about their suitability for new crops and their fertilizer and irrigation requirements. A survey to delineate the soils most suitable for viticulture was also requested by the State Valuation Department (suitability for soil for viticulture being an important determinant of land value) to enable more accurate valuation of land.

The soils of the Padthaway area have already been mapped, but the scale is too small to allow interpretation of the soils within paddocks on farms. A map with this degree of detail was required by both the Extension Officer and the Valuation Department.
It was apparent that mapping was required at two levels, viz. (1) a map to indicate the morphologic types (i.e. 'soil types' identifiable in the field), to be used for general advisory purposes, and (2) a series of maps showing the distribution of specific soil characteristics, and the suitability of soil in different areas for specific crops (suitability determined by combining several soil characteristics important for the particular crop). Such characteristics include depth and texture of topsoil for vegetables, and the water holding capacity of the subsoil layer for vines.

Rapid changes in land use are likely in the area, and we expect periodic requests to re-map the soils using different soil charac-

teristics as criteria. Field data would need to be stored in a way which will allow this re-mapping. Traditional soil survey methods do not give this flexibility as "soil types" are mapped in the field and data is stored in field books. Manipulation of this data is difficult.

In view of these requirements it was decided to use a computer to assist in the storage and manipulation of field data.

Methods

Sampling sites for profile description were located on a grid, with traverses approximately 1 km apart and sampling sites at 100 m intervals along each traverse. Because the geological structure of the area is markedly lineal, the traverses were widely spaced and directed across the lineation.

Prior to going into the field, the traverses were marked on 1: 15 000 aerial photographs. Sampling sites were located in the field by pacing along the path of the traverses. They were then plotted on the photographs.

Detailed profile descriptions were made at each site, profiles being described by observing the cleaned side of a 33 cm diameter auger hole about 1.2 m deep.

At each site data was collected in two parts:
1. *General site details*:
Land form
Micro-topography
Surface drainage
Land use
Native vegetation
2. *Profile description*:
Depth from soil surface to top and bottom of each layer
Colour (including bleach, gley, mottling)
Texture (19 classes)
Stone (type and amount)
Alkaline earth carbonates (tested with 1 Normal HCl)
Fabric (Northcote 1971)
pH (Raupach and Tucker 1959)
Electrical conductivity (1:5 soil:water suspension)
(rarely collected)

After completion of the grid survey, but before computer mapping of the data, 18 back-hoe pits were dug for detailed sampling in representative areas. It was unfortunate that these had to be located prior to computer mapping as the computer could have been used to select representative or 'average' sites. 1350 sites were examined with over 70,000 individual items of information.

Practical details of the data recording and analysis

Recording of field observations

The field recording was designed to be:
- standardized; all profile descriptions to be in the same format with the same characteristics being recorded at each site,
- consistent; each site to be described according to the same criteria (both qualitative and quantitative),
- objective; methods of assessment of soil characteristics, to be objective where possible, rather than subjective,
- numerical or alphabetic in format, to enable direct transfer to punch cards.

To achieve these aims we used a standard recording form (Fig. 1) and code list. Site identification was simple - location plotted on the air photographs were transferred to 1:30 000 Section maps and grid references determined manually from an arbitrary datum.

In coding of site and profile data, simple numerical and alphabetic codes were used to indicate the type, amount, value, etc., of each characteristic. Information about land form, microtopography, surface drainage and land use was collected at all sites, but a record of native vegetation at each site was not mandatory.

The profile descriptions were designed to be complete for every hole and any omissions were regarded as errors (except for electrical conductivity which was only rarely recorded). However, when the amount of stone exceeded 75% no record of colour, texture, fabric of pH was entered.

A requirement for profile classification was for the soil to be described to a depth of 1 m. Therefore, when dense stone prevented drilling of the hole to this depth, the material at the base of the hole was assumed to extend to 1 m. An occassional test hole with a blasting auger supported this assumption.

Field work

In field use we quickly became used to the coding and layout of the cards. With one man observing and the other recording, the average daily performance was 22 holes (including travelling to the site, drilling holes, daily maintenance of equipment, etc.).

The recording forms were checked manually twice by both operators and any errors and omissions detected were corrected. In these cases, new values were inferred from the remainder of the description and by reference to descriptions of nearby profiles.

Data Input and Storage

The data was punched directly to 80-column punch cards: the site details on one card (*header* record) and the profile layer descriptions with one layer per card punched later (*profile* record), given a total of 6 000 punch cards.

PADTHAWAY SOIL SURVEY

Traverse No.
Site No.
Sub-site No.

Land form (position) 1 2 3 4 5 6 7 8 9 10
Micro-topography 1 2 3 4 5 6 7 8 9 10
Surface drainage 1 2 3 4 5 6 7 8 9 10
Land-use 1 2 3 4 5 6 7 8 9 10
Native vegetation 1 2 3 4 5 6 7 8 9 10

Grid reference N.W.
S.W.

Hundred No. . . . .
Section No. . . . .

PROFILE DESCRIPTION

Depth cm.		Colour					Texture		St.		Ca	Fa	pH	E.C.	
		Hue		V/C	GL	BL	M	Des.	No.	No	V				1:5	Ext.

Comments:.....
.....
.....
.....
.....

Classification

Date:

Soil Surveyors:

Fig. 1 Standard field recording form used in Padthaway soil survey

The cards were read into the South Australian Public Service computer (CDC Cyber 73) and stored as card images in an UPDATE library. Storage in this way proved very convenient and allowed ready access to the data for validation, correction and the other manipulations described below.

Data Validation

A FORTRAN program was used to check for omissions and obvious mistakes. Minor errors in field recording could not be uncovered, but detailed checking was designed to reveal gross anomalies. The data was checked for *continuity* of traverse and site numbers, position on the land form and depth records of the layers, *omission* of records, correct gley and bleach *interpretation*, stone type (mainly limestone) and calcium and pH *relationships*, and that the attribute values were within a specified *range* (e.g. pH from 5.0 to 9.5, calcium from 0 to 3%).
We believe that this validation routine uncovered most of the mistakes and omissions. Generally errors occurred at the time of field recording, there being very few punching mistakes.

Data manipulation

The data was first entered as many separate files in an UPDATE library. For each of the 27 traverses there were originally two files in the library, one for header records and the other for profile description records.
Using the profile description data alone three items of information were derived.
1 Primary Profile Form (see Factual Key, Northcote 1971). Three classes were present, Uniform, Gradational and Duplex, there being no Organic profiles in the area. These classes can be strictly determined by applying limits to the texture change from one layer to the next.
2 'Solum depth' (i.e. the 0-100 cm soil layer or the depth of material overlying a layer with greater than 75% stone). Though 75% of stone was generally used, it was varied slightly for some stone types. Obviously this is not a universal description of the solum but it was suitable for this situation. For computer analysis the definition must be explicit; given this, the computer *decision* is objective. Normally in soil surveys a subjective estimation of the solumn depth would be made in the field.
3 Value/Chroma rating, as defined in the Factual Key (Northcote 1971).

The main computer manipulation was classification of the profiles into the main groups recognised in the field. During the field work we noticed about seven fairly obvious and frequently occurring types of soil profiles. It was possible to construct a key to classify the profiles into these types. A FORTRAN program was written to do this and it was successful in that distribution of the soil groups agreed with our recollection of the situation in the field. Changes to the classification were made and the profiles classified

using four slightly different keys. This process is still continuing.
We have plotted these descriptions using the LPPLOT routine. This program places a number identifying the soil class on the printout, the location on the page being determined by the grid reference of the site.

Future Computer Work

Additional work planned for the computer includes:
- production of maps with single specific factors, e.g. solum depth, surface texture, depth and thickness of clay layers, etc.,
- cluster analysis to select 'natural' profile groups,
- calculation of variability within and between groups.

Problems and successes

The use of recording form in the field was completely successful. It made recording quicker and easier, and the cards themselves are a very useful method for storage and rapid manual retrieval of soil descriptions.
The greatest problem has been the long delay in computer processing. This work has been limited by our lack of expertise in programming and use of the computer. Being located 350 km from the computer has also slowed progress. To overcome these problems a computer specialist should be part of the soil survey team. It would be desirable for him to be versed in soil description and techniques of soil surveying.
The classification technique used has objective and subjective parts. Obviously the programme for assigning the profiles to the classes is subjective - it is based on selected variables. However the *process* of assigning a profile to a particular class is completely objective in contrast to usual field techniques. It is a consistent process in that it can be repeated and will always give the same results.
In view of the subjectivity of the classification technique we hope to use cluster analysis to group the profiles. In this way it may be possible to identify 'natural' groups which we know occur in the field. This should not suffer the particular subjectivity present in the method used so far, although cluster analysis also involves subjectivity in the selection of soil characteristics.
The method used obvious advantages over the traditional way in which detailed soil surveys are carried out. Firstly the soils are not classified until all sites have been examined. This replaces the selection of 'soil types' from a reconnaissance survey prior to the main grid survey. Inconsistency and bias in the classification is therefore avoided, as is the problem of classifying new profiles which may be discovered during the survey and which do not fit into soil groups already defined.
The very large amount of grid data involved in detailed surveys has previously been a stumbling block in classifying soils after the survey has been completed - the computer now makes this possi-

ble. In addition, the classification can be changed many times using different criteria as it is not constrained by decisions made at the start of field work.

We have not yet calculated the variability of various characteristics within the soil groups but the computation will present no difficulties. The unbiased site selection makes calculations statistically valid. Though one may not be able to use a measure of variability such as standard deviation, it will be possible to construct histograms and express the variability in terms of percentages of sites within specified ranges of each variable.

References

Northcote, K.H. (1971). A factual key for the recognition of Australian Soils, 3rd Edition. Rellim Press, Adelaide, South Australia.

Raupach, M. and B.M. Tucker (1959). The field determination of soil reaction, J. Aust. Inst. Agric. Sci. 25: 129-133.

Storage and retrieval of soil profile classification and morphological data

K.M. Stackhouse, Department of Agriculture, Lanceston South, Tasmania

Introduction

One set of information acquired during a soil or land system survey consists of descriptions of soil profiles. Values of several attributes of soil samples taken down the profile are recorded, e.g., texture, colour, condition of surface soil, pedality, structure, fabric, consistence, pH, boundaries, designation and strength of horizons. Soils may be named as a soil type or classified according to Nothcote's Factual Key (Northcote 1971) and/or some other system. Further details of location, geology, land form, vegetation etc. are usually recorded at each profile site.
The quantity of this type of information collected during the survey is such that usually only a fraction appears in the published survey, the bulk possibly remaining in the surveyors field notebook or other document that is not readily accessible.
Beckett and Burrough (1971) suggest that the following questions might be asked by users of a soil survey:

1. a What classes of soil are present?
   b In what proportions do they occur?
   c What proportions of the area are occupied by soils with particular properties or particular ranges of one or more properties?
2. a What is the soil class at any site of interest in the area?
   b What are the properties of the soil at any site in the area?
3. a Where can soils of a particular class be found in the area?
   b Where can soils of particular properties be found in the area?

The soil map and descriptions of the few published profile descriptions may not be sufficient to answer to the user's satisfaction one or more of these questions.
The following describes a system of recording information on soil profiles on magnetic tape and for retrieval of parts of the information relevant to a user's particular requirements.
The computer programs have been written in COBOL for a CDC3200 computer operating under MSOS. COBOL, the language commonly used for processing business data, was used because the data consists mainly of non-numeric items and processing involves manipulation of large files of information.

Input

One method that can be used for data input is to use a predetermined numeric coding for the values of the various attributes

of the soil samples. This system is inflexible and cases may arise during use of the code where some unusual value of some attribute has not been previously defined. I prefer to encode, as fas as is practicable, the original non-numeric values. Of course, where retrieval according to a specific value of some attribute is required, standardisation of the non-numeric values of the attributes is necessary. If the records are retained in a numeric code, a user of the file of information needs to know:
1. The coding system for the attributes
2. The structure of the records
If on the other hand the records consist of the actual values of the attributes as line printer images as shown in table 1, retrieval of information may be obtained by using a program containing condition statements querying the actual value of a particular attribute of interest rather than its coded value.
In COBOL the structures of records are described in the DATA DIVISION of the program in contrast to FORTRAN where the FORMAT statements define the structure.

For instance in FORTRAN we might have:

```
READ (60,2) FARM, DIST, SOIL, TEXT
2 FORMAT (4 A 8)
READ (60,4) PARENT, CLASS EAST, ORTH
4 FORMAT (2 A 8, 2 F 6.0)
```

In COBOL these 4 statements would require:

```
DATA DIVISION.
FILE SECTION.
FD CARD-FILE LABEL RECORD IS OMITTED
      DATA RECORD IS KARD.
01 KARD PIC X(80).
WORKING-STORAGE SECTION.
01 FIRST-REC.
      02  FARMER          PIC X(8).
      02  DISTRICT        PIC X(8).
      02  SOIL-CLASS      PIC X(8).
      02  TEXTURE         PIC X(8).
01 SECOND-REC.
      02  PARENT-MATERIAL   PIC X(8).
      02  CLASSIFICATION    PIC X(8).
      02  EAST-COORD        PIC 9(6).
      02  NORTH-COORD       PIC 9(6).
PROCEDURE DIVISION
      OPEN INPUT CARD-FILE.
      READ CARD-FILE INTO FIRST-REC AT
      END GO TO.....
      READ CARD-FILE INTO SECOND-REC AT
      END GO TO.....
```

Although the coding for COBOL is much more tedious than for FORTRAN, the structure of records (entry names may have up to 30 characters) is more obvious than from the FORMAT declarations in FORTRAN.

In my system initially input was by means of punched cards. For one soil profile the input records consist of the following. Two cards (1 and 2) contained the information shown in the first two lines of Table 1. For each soil sample examined in a profile, three cards (3, 4, 5) contained the information shown in Table 1 for a sample. When encoding data from a large number of profiles, I found it more prudent to allow punch operators to punch cards in sequence of all no. 1, all no. 2, all no. 3, etc. and then rearrange by hand sorting into the sequence 1, 2, 3, 4, 5, 3, 4, 5, etc. Acquisition of a tape encoding system by the computer centre required changing input to tape. A tape is encoded in the same sequence as was used for cards and a SORT programme is now used to produce an input tape with the desired sequence of records.

Output

The operating system of the computer centre used for this project requires that large amounts of output be written on magnetic tape which is transferred to a subsidiary system consisting of a CDC 160A computer, tape drive and line printer for listing the output. Table 1 shows an example of the output following processing of the input tape by a programme SOIL-OUT.
Table 2 shows an example of the output following processing of the SOIL-OUT output tape by a programme SOIL-SUMMARY.
Updated files of both types of records are maintained.

Retrieval

Either of the two files may be interrogated for information pertinent to the requirements of a user.
For instance a simple programme SOIL-SEEK can be used to print-out from the file of Table 2 records all profiles:

(i) Within a given rectangle of co-ordinates
(ii) With common Factual Key symbols
e.g. all D or all Dy or all Dy3 or all Dy 3.2
(iii) A combination of (i) and (ii)

Antoher simple programme PRINT-PRO may be used to provide output of a required set of whole profile records from input of profile identification numbers.

Rather than use a general purpose programma for retrieval of those profiles with a particular value of some attribute I prefer to modify the coding of a search programme for the particular need.

For example suppose I wish to search for profiles with the attribute of a gleyed subsoil. Table 1 indicates the word GLEY first occurring at line 33. I wish to record the first two lines of the record, degree of GLEY and the depth at which this condition first occurs. The outline of the COBOL coding required is:

Table 1 *Record of a soil profile description*

1	PROFILE NO.00108		497050YE 869380YN	FYFE I		PADDOCK 33		SITE 4 CRESSY			72-17- 9		RCK
2													
3		UG6.22	TERTIARY SEDIMENTS KINBURN ASSOCIATION										
4													
5	CM.TEXTURE	.COARSE FRACTION	.COLOURS		V/C	.CONDITION	.PED ALITY	.STRUCTURE	.FABRIC	.CONSISTENCE	.PH	.CARBONATE	.BOUNDARIES	.HORIZON
8	008.MEDIUM CLAY	.SOME FE GRAVEL	.10 YR3/2.		1	.HARD SETTING.	.PEDAL.	.COHERENT .MODERATE .MEDIUM .GRANULAR	.ROUGH	.FRIABLE .DEEP .CRACKS	.5.5		.DIFFUSE	A1 .STRONG
13		SALTS 0.025 CHANNEL STAINS												
15	018.MEDIUM CLAY	.SOME FE GRAVEL	.10 YR4/1.		2		.PEDAL.	.COHERENT .MODERATE .MEDIUM .GRANULAR	.ROUGH	.FRIABLE	.5.9		.ABRUPT	A1
20		SALTS 0.01 CHANNEL STAINS												
22	028.MEDIUM CLAY	.SOME FE GRAVEL	.10 YR4/2.		2		.PEDAL.	.COHERENT .WEAK .FINE .GRANULAR	.ROUGH	.FIRM	.6.6		.CLEAR	AB
27		SALTS 0.016 SANDY												
29	042.HEAVY CLAY	.SOME FE GRAVEL	. 5 Y 5/1. .10 YR5/8.STRONG MOTL. .2.5YR5/8. . 5 Y 5/1.STRONG GLEY.		2		.PEDAL.	.COHERENT .WEAK .FINE .BLOCKY	.ROUGH	.FIRM	.6.8		.DIFFUSE	B1
34		SALTS 0.023												
36	050.HEAVY CLAY	.SOME FE GRAVEL	. 5 Y 4/1. .10 YR5/8.STRONG MOTL. .2.5YR5/8. . 5 Y 4/1.STRONG GLEY.		2		.PEDAL.	.COHERENT .WEAK .FINE .BLOCKY	.ROUGH	.FIRM	.6.6		.DIFFUSE	B2
41		SALTS 0.04 BRITTLE												

Table 2 Location and classification of soil profiles

DEP OF AGRIC SOIL SUMMARY

PROFILE NO.00091	00501831ME05382548MN	KEAM P E	PADDOCK 18	SITE	12 CRESSY	72- 7- 9	RCK
DY3.82	QUAT BRUMBY TERRACE BRUMBY ASSOCIATION						
PROFILE NO.00092	00502001ME05382824MN	KEAM P E	PADDOCK 18	SITE	13 CRESSY	72- 7- 9	RCK
DY3.12	QUAT BRUMBY TERRACE BRUMBY ASSOCIATION						
PROFILE NO.00093	00501955ME05382833MN	KEAM P E	PADDOCK 18	SITE	14 CRESSY	72- 7- 9	RCK
DY3.12	QUAT BRUMBY TERRACE BRUMBY ASSOCIATION						
PROFILE NO.00094	00501919ME05382841MN	KEAM P E	PADDOCK 18	SITE	15 CRESSY	72- 7- 9	RCK
DY3.12	QUAT BRUMBY TERRACE BRUMBY ASSOCIATION						
PROFILE NO.00095	00501873ME05382850MN	KEAM P E	PADDOCK 18	SITE	16 CRESSY	72- 7- 9	RCK
DY3.12	QUAT BRUMBY TERRACE BRUMBY ASSOCIATION						
PROFILE NO.00096	00505165ME05387430MN	SKIRVING K	PADDOCK 9	SITE	1 CRESSY	72-13- 9	RCK
DD2.21	TERTIARY SEDIMENTS KINBURN ASSOCIATION						
PROFILE NO. 00097	00505165ME05387430MN	SKIRVING K	PADDOCK 9	SITE	2 CRESSY	72-13- 9	RCK
DG1.61	TERTIARY SEDIMENTS KINBURN ASSOCIATION						
PROFILE NO.00098	00505165ME05387430MN	SKIRVING K	PADDOCK 9	SITE	3 CRESSY	72-13- 9	RCK
DG2.21	TERTIARY SEDIMENTS KINBURN ASSOCIATION						
PROFILE NO.00099	00505165ME05387430MN	SKIRVING K	PADDOCK 9	SITE	4 CRESSY	72-13- 9	RCK
DY3.42	TERTIARY SEDIMENTS KINBURN ASSOCIATION SANDY SURFACE						
PROFILE NO.00100	00505165ME05387430MN	SKIRVING K	PADDOCK 9	SITE	5 CRESSY	72-13- 9	RCK
DY3.51	TERTIARY SEDIMENTS KINBURN ASSOCIATION SANDY SURFACE						
PROFILE NO.00101	00505165ME05387430MN	SKIRVING K	PADDOCK 9	SITE	6 CRESSY	72-13- 9	RCK
DY3.42	TERTIARY SEDIMENTS KINBURN ASSOCIATION SANDY SURFACE						
PROFILE NO.00102	00505165ME05387430MN	SKIRVING K	PADDOCK 9	SITE	7 CRESSY	72-13- 9	RCK
DY3.22	TERTIARY SEDIMENTS KINBURN ASSOCIATION SANDY SURFACE						
PROFILE NO. 00103	00505165ME05387430MN	SKIRVING K	PADDOCK 9	SITE	8 CRESSY	72-13- 9	RCK
DY3.82	TERTIARY SEDIMENTS KINBURN ASSOCIATION SANDY SURFACE						
PROFILE NO. 00104	00505165ME05387430MN	SKIRVING K	PADDOCK 9	SITE	9 CRESSY	72-13- 9	RCK
DG2.42	TERTIARY SEDIMENTS KINBURN ASSOCIATION SANDY SURFACE						

```
FILE SECTION.
FD FILE-REC ..... DATA RECORD IS SOIL.
01 SOIL PIC X(136).
FD OUT-REC ..... DATA RECORD IS OUT.
01 OUT PIC X(136).
WORKING-STORAGE SECTION.
      01  LINE-1.
              02  CHAR-1  PIC X.
              02  REMAIN  PIC X(135).
      01  LINE-2   PIC X(136).
      01  SEARCH-LINE.
              02  CHARAC-1  PIC X.
              02  DEPTH     PIC X(3).
              02  FILLER    PIC X(27).
              02  GLEY.
                  03  STRENGTH   PIC X(6).
                  03  FILLER     PIC X.
                  03  RWORD      PIC X(4).
              02  REST  PIC X(94).
      01  DEPTH-GLEY  PIC X(3).

PROCEDURE DIVISION

      P1.     READ FILE-REC INTO LINE-1  ....
              IF CHAR-1 NOT = '1' GO TO P1.
      P2.     READ FILE-REC INTO LINE-2  ....
      P3.     READ FILE-REC INTO SEARCH-LINE ....
              IF CHARAC-1 = '1' MOVE SEARCH-LINE
              TO LINE-1 ELSE GO TO P4.
              GO TO P2.
      P4.     IF DEPTH IS NUMERIC MOVE DEPTH TO DEPTH-GLEY.
              IF RWORD = 'GLEY' GO TO P6
              ELSE GO TO P3.
      P6.     MOVE 'O' TO CHARC-1.
              MOVE DEPTH-GLEY TO DEPTH.
              WRITE OUT FROM LINE-1.
              WRITE OUT FROM LINE-2.
              WRITE OUT FROM SEARCH-LINE.
              GO TO P1.
```

Output consists of the first two lines of table 1 with a further line containing 042 ..... STRONG GLEY.
Similarly simple programs can be written for searching for profiles wherein other selected attributes have particular values.

Discussion

Computing of the Factual Key classification of the profile from the values of the attributes of the samples has not been attempted, although there is no basic logical problem in assigning a profile to the appropriate class.
Profile records within the file could be extended to include chemical data of soil samples and other data. Different workers will

without doubt require different sorts of information about soils to be recorded and retrieved. Thus a universally accepted system even among Australian workers is unlikely to eventuate. I am suggesting that provided the structure of the files of this type of information are known, a user may write programmes to retrieve the sort of information he requires from a file.
Personally I have found the system I have adopted to be useful in providing information on Tasmanian soils to agronomists. I also feel that COBOL has advantages over FORTRAN and possibly systems such as INFOL 2 (Pummeroy 1973) for processing this type of data.

References

Beckett, P.H.T. and P.A. Burrough (1971). The relation between cost and utility in soil survey. J. Soil Sci. 22: 466-489.
Northcote, K.H. (1971). A Factual Key for the Recognition of Australian Soils. Rellim Tech. Publs, S. Australia.
Pummeroy, N.R. (1973). INFOL 2 Reference Manual. Tech. Note No. 40. CSIRO Div. Comp. Res., Canberra.

Data delineation and computer techniques for line printer mapping and tabulation

R.S. Cormack, Department of Primary Industries, Brisbane, Queensland

Introduction

As the Queensland Department of Primary Industries becomes more involved in land resource studies new procedures and techniques are being sought for processing the recorded data. The advent of computer systems has provided a means for quickly and efficiently sifting through masses of data which previously had accumalated without being fully analysed. If computer systems are to be utilised then the manner in which data is collected and recorded must be examined in relation to input, manipulation, and display from the computer.

When maps are to be displayed by computer a distinction has to be made between point and area data. Point data relates to a site while area data refers to a delineated area of land. Map boundaries can be generated or extrapolated from point data using pattern analysis and/or contouring techniques (Webster and Burrough 1972). Maps may be output from area data by displaying symbols within the boundaries delineated.

To date three land resource studies have been completed in Queensland using a computer to manipulate and display area data in both map and tabular form. This paper discusses our concepts for the delineation and processing of area data for line-printer mapping by computer. The processing of point data is also discussed.

Delineation of area data

One of the most difficult aspects of mapping is the delineation of boundaries. This is particularly so when a large number of attribute classes which are not separated by the same boundary are examined simultaneously. The problem is to generate a map of positional information for the study region so that boundaries do separate one or more of the described attributes. Computer processing gives rise to an additional problem, viz. to arrange some form of internal representation of the positional map within the computer.

For the purpose of this paper a 'map unit' is defined as a discrete region in which the recorded attributes are considered reasonably homogeneous. To prepare a positional map involves delineating, on a map (or otherwise) of the study region, the component

map units.

Two procedures have been used for generating and recording the positional information. The first is to prepare a line map of the study region which separates the map units. The second is to impose a regular recording unit, such as a cell, over the study region.

1. Line map

Preparation of positional map

There are several ways in which multi-attribute maps can be prepared. One procedure is to overlay the individual attribute maps onto a single map. The resulting regions on the single map then become the map units. For the Moreton Region Non Urban Land Suitability Study (MRS) (Anon. 1974) a soil map prepared by the CSIRO Division of Soils was selected as a basis for the delineation of the map units. With the aid of this map, patterns of soil, topography and vegetation associations were delineated onto a photo-mosaic of the Moreton Region. These patterns were further sub-divided into the map units on the basis of catchment, existing land use, land slope and clearing. The map unit boundaries were then transferred onto a cartographic map base of the Moreton Region.

Computer representation

Continuous line map boundaries require conversion to a digital representation before they can be stored in a computer. The technique we adopted was to break the continuous line boundaries into a series of cell increments for computer storage as cell or matrix data.
This can be done automatically by the use of a digitizer, or encoded manually by overlaying the base map with a regular grid. The cell size is related to available computer output devices. If matrix printers, plotters or microfilm devices are available, small cell sizes are possible resulting in almost smooth line maps. For line-printer output the cell sizes correspond to the physical size of line-printer characters. A regular grid corresponding to the physical dimensions of the line-printer characters was overlayed on the base map used in the MRS. The positional map information was encoded manually for computer input and storage.
Computer processing of line maps may necessitate the creation of two computer files. The computer 'map' file contains the digital positional information of the map units. The computer 'attribute' file contains the attribute information for each of the map units. These two files are processed together to generate and display desired maps.
A disadvantage of line-printer techniques is the inability to vary map scale, other than by multiples of two from the encoded base map. This disadvantage may be overcome by using a plotter to map at the intermediate scales. Matrix printers and microfilm devices do allow propertional variation in cell size and also have shading capabilities which are desirable for map display.

2. Regular Recording Unit

The disadvantage of the technique used in the MRS was the manual preparation of the positional map and manual digitization before input to a computer. To overcome these problems a total grid cell procedure (Milton and Rosenthal pers. comm.) GORB (Generation of Resource Boundaries) was developed for the North Pine Dam Catchment Area Land Use Study (Milton, James, Briggs 1975). This technique is subsequently being used in the upper Burdekin catchment survey which is part of joint State/Commonwealth study of the Burdekin River Basin. GORB has also been adopted by the Soil Conservation Service of New South Wales.

Preparation of positional map

GORB is a total grid-cell procedure in that data is recorded on a regular cell basis as well as being stored in the computer as cell or matrix data. Preparation of the positional map merely involves overlaying the study region with a regular grid. Each cell is examined in turn and within each a number of map units are described but not delineated. Data for each map unit is collected on a statistical basis as a proportion of the cell. The size of each cell and the maximum number of units that may be described within each cell is dependent on the level of the survey. GORB thus provides a mechanism for recording the dominant as well as the sub-dominant information for a cell. Sub-dominant recording ensures that variations are not lost due to the placement of the cells. With map display GORB does not, or could not, attempt to draw precise boundaries but indicates that boundaries exist.

Computer representation

The positional information is in a digital format at the recording stage and thus suitable for computer processing. This has an added advantage that once data for a study sub-region has been recorded it can be processed before waiting for recording of the entire study region to be completed.
Computer processing of regular recording units will normally require the establishment of one computer file only as the positional information can be recorded as an attribute.

Data scale

Computers provide the ability to display maps at any desired scale. If a map scale is to vary so must the detail of the presented information. When recording raw data an attempt should be made to associate the data with an anticipated display scale.
With the GORB approach, the surveyor when recording raw data, is forced to look at cells of land whose area corresponds to that represented by a line-printer character at the proposed mapping scale. The resource classes to be described within the cell may occupy a minimum of 1% of the cell. The minimum, however, is seldom less than 10% with an imposed maximum of six resource classes per

cell. To vary the detail of the data simply involves varying cell size and describing up to the same number of resource classes per cell. As GORB imposes a systematic recording procedure on the surveyor the entire study region is looked at the same level of detail.

Line-printer mapping at a scale of 1:100 000 required the initial division of a study region into 8 ha cells. Although the location of the resource class data is restricted to somewhere in the cell its presence can be indicated. Hence, in this case, map units of less than eight hectares will be recorded and map display can indicate their presence and locate their position roughly by the cell location.
A cell is divisable into four equal cells thus providing a mechanism for variable scale (variable detail) recording and display. However, the mapping scales are restricted and belong to a set and division of a cell doubles the mapping scale. To achieve variable detail recording requires the largest cells (small amount of detail) for the purpose of the survey to be overlayed on the study region. These cells are divided where more detail is required.

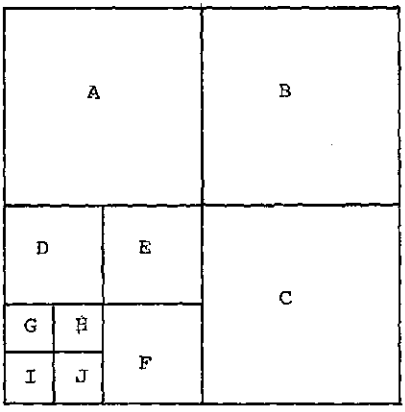


Fig. 1. Cell division

For example in Fig. 1 data for cells A, B, C, is recorded at 1:100 000. Data for cells D, E, F is recorded at 1:50 000. Data for cells J, H, G, I is recorded at 1:25 000.
To map at a small scale involves summing data for all large-scale cells, covered by a single small-scale cell, and displaying the data as a single cell. To do the reverse involves displaying as blank the large-scale cells covered by a single small-scale cell. From Fig. 1 to map at a scale of 1:100 000 would involve displaying cells A, B, C as they exist and summing the data over cells F, G, H, I, J, D, E to display as a single cell. To map at a scale of 1:25 000 involves displaying cells J, G, H, I as they exist and blank out the cells covered by A, B, C, D, E, F.
If all cell co-ordinates (bottom left-hand corner) are based on the eastings and northings of the Australian Map Grid, and can thus

be related to a common point for each study region, then the data from the different regions can be combined without cells overlapping. This will allow piecemeal generation of a state or national data bank from regional surveys.

Computer processing and output

A computer has been used to perform two types of processing on our data. The first and simpler of the two is a direct extraction and presentation of the data in a reduced form. The second involves a manipulation of the data, to generate information, and display the result. With this type a prior derivation of a model is usual. With both types of processing the output is displayed in either map or tabular form.

Direct Extraction

Tabular outputs are sorted listings of the recorded data. They interrelate the attribute classes and indicate the magnitude of the areas involved. Attributes to be extracted are chosen singly or in combination. When combinations of attributes are required breakdown for attribute sort order is specified by the user.
A general tabulating program has been written for each survey. The user supplies the codes for the desired attributes and their classes as input to the program. The resource classes and areas of land involved are automatically listed with each group of attribute classes.

Map outputs distinguish between the attribute classes or combination of attribute classes by representing each class or combination with a unique line-printer character. Each map is labelled with a legend identifying the line-printer character with its representation. A user provides the retrieval criteria to the mapping program in the same manner as the tabulating program. Each time a new class or combination is detected the program allocates a character from a symbol array. Fig. 2 is a display of the stored agricultural capability map for the MRS.
With the GORB approach, where there may be up to six resource classes per cell, the dominant (or other) resource class is selected and mapped. GORB is orientated towards producing intensity maps where a character designates the proportion of a cell occupied by a class or combination of classes.

Manipulative Extraction

Manipulative extraction of data requires some logical operation to be performed on the data. Relationships between attributes are defined by the user prior to retrieval. The relationships are encoded into the computer which manipulates the data in a defined manner and displays results in either map or tabular form.

Our procedure has been to write a general program for extraction and manipulation and to provide output in either map or tabular

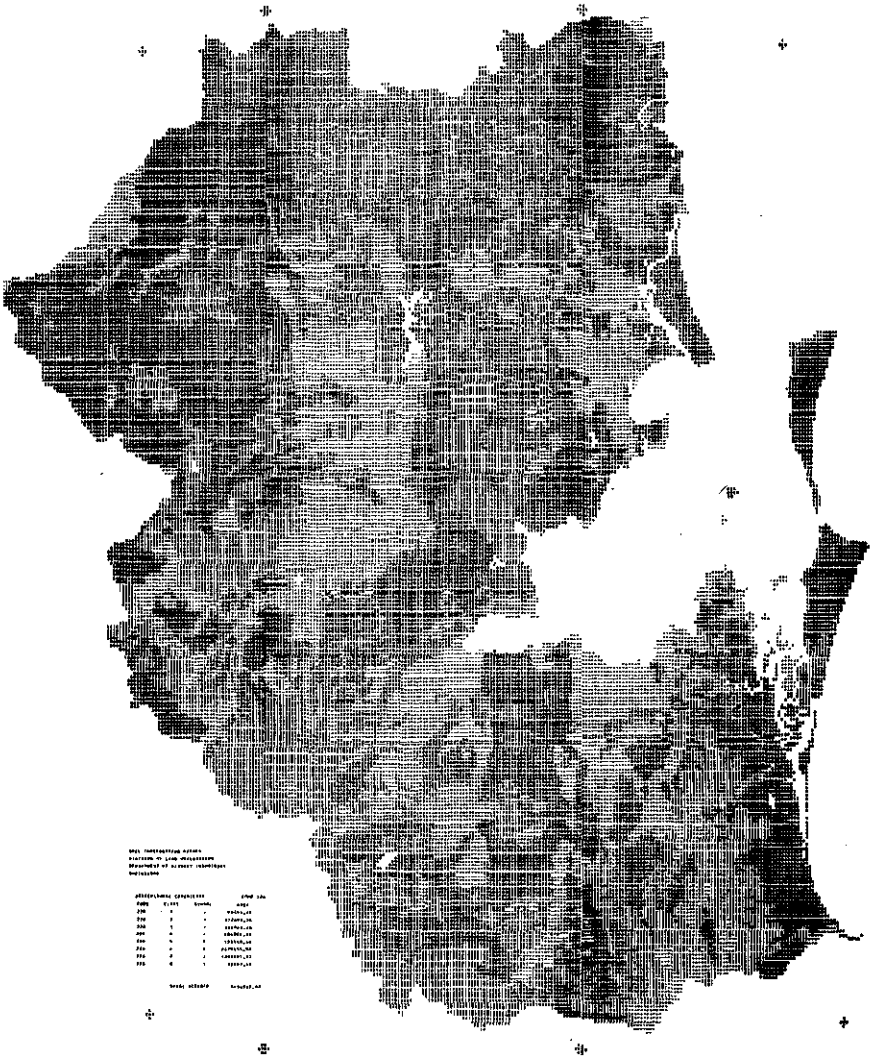


Fig. 2 Display of the stored Agricultural Capability map for the Moreton Region Study. Crosses outside the map boundary are for the registration of a transparent overlay.

form or both. Up to ten sets of selection criteria are provided by the user as input. These criteria enable Boolean 'and' relationships between attributes and Boolean 'or' relationships between attribute classes to be tested. When a set of selection criteria is satisfied control is transferred to a user provided routine which can test for more involved logical relationships and perform numerical operations. Each set of selection criteria qualify the data for an individual purpose. Each purpose may be discrete, such as cotton growing capability,or belong to a set such as one of five pastoral capability classes.
For each set of selection criteria an intermediate file is created which contains one entry for each map unit that satisfies the criteria. For tabular output a table is produced from each intermediate file. This table contains the map unit entries together with the area of land involved.
Intensity maps and dominant maps are displayed from the intermediate files. Intensity maps are output from each intermediate file. The line-printer characters or symbols of the map represent the proportion or cell allocated to a purpose. A single dominant map is generated from all the intermediate files. The character symbols of a dominant map represent the purpose or enterprise that occupies the majority of the cell.

Fig. 3 is an intensity map representing the generated pastoral capability class 2 for the Upper Burdekin Basin Study.

Map Presentation

Map presentation can be enhanced by preparing a transparent cartographic overlay with the major towns, physical features, etc, marked on it. This transparency is then overlaid on the computer map output. The crosses on Fig. 3 outside the map region are for the registration of the transparent overlay.

Point site data

Techniques are being looked at for computer-processing of point site and profile data for both broad-scale and detailed studies. In the Southbrook region of the Darling Downs point site and profile data is being sampled and recorded on a 250 metre grid. The recorded data is being assembled into a sequential computer file. Envisaged computer processing of this data will entail point site plotting, contouring, statistical analysis, pattern analysis and sorting. Software packages are available that will perform these types of computer processing. However, the limiting problem with previous surveys of this type has been the extraction of data from the computer files in a format suitable for input to the required application program.
The approach with this survey has been to develop a computer program which will interface the data file to the application program via an intermediate file. As most requests of the data file simply require sorted listing of the recorded data the program has built-in sorting facilities.

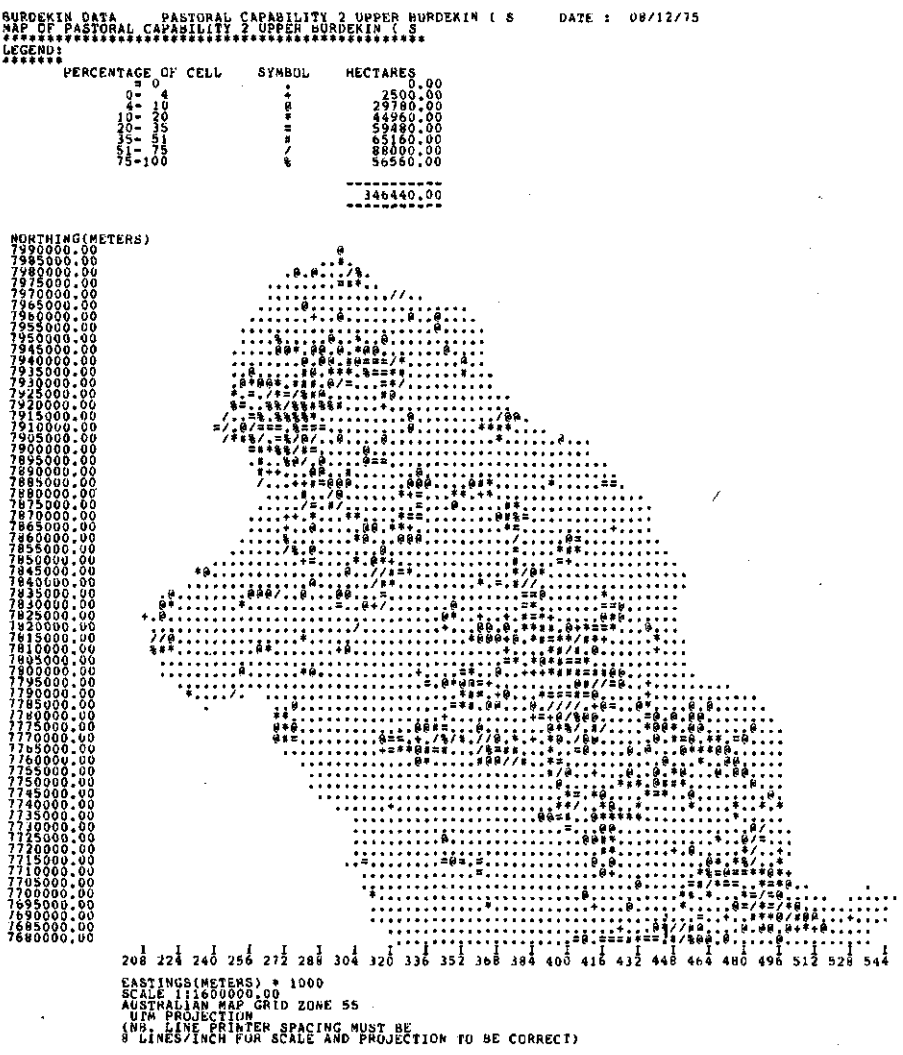


Fig. 3 Intensity map of the generated pastoral capability 2 Upper Burdekin Study.

Conclusion

Computer processing of land resource data is not only a computing problem. Area data can be recorded from line maps or by the imposition of a regular recording unit on the study region. Although line-printer techniques do not provide for the display of concentional thematic maps they do display the same information. However, map information is only meaningful if some attempt is made to justify the correctness of data at the mapped scale. The GORB approach has proved its ability to provide correct and timely information.
As computer systems develop physical processing constraints change. Parallel with this are changing attitudes to the processing and content of recorded data. Information processing procedures must be amenable to change and are thus on-going and developing.

References

Webster, R. & P.A. Burrough (1972). Computer based soil mapping of small areas from sample data. J. Soil Sci. 23:210-221.

Anon (1974). Moreton Region Non Urban Land Suitability Study. Dept. of Primary Industries, Brisbane, Qd.

Milton, L.E., S. McF. James & H.S. Briggs (1975). A study of land resources hazard and management for the catchment of North Pine Dam. Appendix D, North Pine Dam Catchment area land use study. Dept. of Local Govt., Brisbane, Qd. 1975.

The National Soil Fertility Data Bank and methods for data manipulation

J.D. Colwell. Division of Soils, CSIRO, Canberra, A.C.T.

Abstract

The National Soil Fertility Data Bank has been complied to provide bases for studies of relationships between soil fertility, fertilizer requirements, soil composition, type of soil and climate. Such studies require many and varied selections of data and experimentation in the selection of appropriate statistical models, and this has posed a considerable practical problem in the use of the bank.

Computers subroutines CORA and SAM have been devised t6 facilitate studies on correlation and regression relationships. These subroutines and their use are described and computer listings of the programs and examples of their use are given on microfiche film.

Introduction

The National Soil Fertility (N.S.F.) Data Bank has been compiled from data obtained with the National Soil Fertility Project (Hallsworth, 1969; Colwell, 1976) and consists of extensive sets of data representing various aspects of the fertility, composition and morphology of important soils across the southern portion of Australia. The bank was established to provide bases for studies on relationships between selections of these variables, on a nation-wide scale, covering for example the relationships between fertilizer requirements, types of soil, soil composition and locality, and interrelationships between different soil analyses (Colwell 1970, 1971, 1977b). The use of the bank for these purposes leads however to practical problems due to the many selections of data that have to be made to represent different soils and localities, the many combinations of data variables that have to be tried in establishing and assessing relationships, and incompleteness and inconsistencies amongst the data sets. The major obstacle to the use of the bank for the purposes for which it was intended, has been in fact the labour required of the user in obtaining the data he wants from the bank and in then transferring them to appropriate statistical-economic computer programs. This type of difficulty seems basic to the use of such banks, particularly when data are for many related variables and have been provided by many workers distributed amongst a variety of research organisations, as for the N.S.F. Project. The difficulty has been met for the N.S.F. Data Bank by the development of data selecting subroutines for use in conjunction with standard statistical-economic subroutines. Two of these subroutines are described in this paper to illustrate the

methods used in studies with this particular bank; listing of programs and examples of output from them are presented on the microfiche stored in the pocket inside the back cover of this publication. A similar series of subroutines (GEN series) are described by Colwell (1977a) for studies on regional fertilizer requirements.

The selection problem

The data for the N.S.F. Data Bank were obtained from a series of fertilizer experiments that were carried out in a 5-year period on sites representing particular soils and localities. Yield data from the wheat or pasture used for these experiments were used to calculate values representing different aspects of the fertility at the sites, thus providing the soil fertility data. Soil samples representing the soil profile at the sites were analysed by a range of standard procedures and used to calculate profile trends, providing soil data. In addition, climatic data were recorded for the sites, agronomic features noted, and so on, to provide data on other factors likely to affect the fertility of the site. In this way data for about 370 variables were recorded for each of about 600 sites. Although in theory this all involved straightforward field and laboratory work directed towards the establishment of the bank, in practice difficulties arise resulting in incomplete sets of data for many of the variables. Misunderstandings concerning sampling and analytical procedures arose amongst the many workers, over the wide area of the project (3000 km from east to west), accidents occurred resulting in losses of data, and some centres were not able to fulfil the large amount of work required of them. The final data bank consequently contains many gaps as represented diagrammatically in Fig. 1.

Site	Variable					
	V_1	V_2	V_3	V_4	V_5	$V_6 \cdots$
1	-	-			-	
2	-	-			-	-
3	-	-	-		-	-
4	-	-	-		-	-
5			-	-		-
6	-	-	-	-		-
7	-	-	-	-		-
8	-	-	-	-		
9	-	-	-	-		-
10	-	-	-	-		-
11	-	-	-		-	-
...						

Fig. 1

The study of relationships with the data requires the selection of groups of variables (columns in Fig. 1) to represent particular aspects of the fertility, soil composition etc., and of sites (rows in Fig. 1) to represent particular soils and localities, and to avoid missing data. This selecting of data, in effect by row and

column from a large data array, must be performed hundreds of times over for statistical studies on relationships, the selected data in each instance being processed by a standard computer program, possibly with mathematical transformations for homogeneity of variances ot to allow for curvilinear relationships. This selection problem has led to the development of the CORA, SAM and associated subroutines now described.

Subroutines CORA *and* CORREL

The subroutines CORA selects data from an array V on the basis of integer references to row and column of each datum, and computes correlation coefficients by the subroutine CORREL for all combination pairs of the selected variables. The mean, standard deviation, range and coefficient of variation are also computed for each variable. To use CORA, a main program is written to read the appropriate data into the arrays AME, V, NR and NC declared in a common statement (see COMMON statements in microfiche listings), where AME contains identifying names, as alphanumeric data for each of the selected variables, V contains the selected data arranged in columns corresponding to the names in AME, NC contains integers identifying columns of data in V to be selected and NR similarly contains integers identifying rows of V to be selected for correlation computations. The program should read or specify data and then initiate the computations by teh calling statement

```
CALL CORA (n, k)
```

where parameter n is the number of values for each variable i.e. number of integers in NC, and parameter k is the number of variables i.e. of integers in NR.

The computation of correlation coefficients etc. is carried out by CORREL and tests of significance by PRBF (based on the function PRBF given by Veldman, 1967). These subroutines may be used independently of CORA by simply reading data into AME and V of the common statement of CORREL (see microfiche listing) followed by the calling statement

```
CALL CORREL (n, k)
```

Such direct use of CORREL precludes however the possibility of data selections provided by the column and row specifications of CORA.
The capacities of CORA and CORREL are limited only by the dimension specifications for the respective common statements. The specifications given on the microfiche listings are respectively,

```
COMMON W(200,90)AM(90),NC(90), NR(200,90),AME(90)
```

and,

```
COMMON V(200,90), AM(90)
```

These allow the reading and selection of up to 200 data values for each of 90 variables. This capacity has been found ample for all the computations required so far with the data bank, and is made possible by the extended storage facility provided by the LEVEL 2 declaration for the CDC Cyber 76 computer. Smaller dimension declarations may be necessary for other computers.

The subroutines CORA, CORREL and PRBF are listed on the accompanying microfiche film, together with an example main program CORS. A portion of the data bank read by CORS is also given, followed by integer values specifying rows and columns to be selected for the correlation computations. The examples produce large arrays of correlation coefficients (90 x 90) for various selections of the data and some of the outputs are also given to illustrate the uses of these subroutines.

Subroutines SAM *and* REG

The subroutine SAM (Select And Model) also selects data from an array V, by row and column as with CORA, but for the computation of regression coefficients and the associated analysis of variance. It also arranges and transforms the selected variables for linear regression models of the following general forms:

ML	Model
1	$Y = b_0 + b_1X_1 + b_2X_2 + \ldots$
2	$Y = b_0 + b_1X_1 + b_2X_1^2 + b_3X_2 + b_4X_3 + \ldots$
3	$Y = b_0 + b_1X_1 + b_2X_2 + b\,X_1^2 + b_4X_2^2 + b_5X_3 + \ldots$
4	$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2 + b_4X_1^2 + b_5X_2^2 + b_6X_3\ldots$
5	$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2 + b_4X_3 + \ldots$

where the dependent variable Y, and the independents X_1, X_2 ... may be any of the variables stored as columns in V. The form of the models may be further modified by transforming the selected data columns, before their substitution in these models, as follows:

LC	TRANSFORMATION OF V_j	
1	V_j	(no transformation)
2	Log V_j	(natural logarithm)
3	e^{Vj}	(exponential)
4	$\sqrt{V_j}$	(square root)
5	V_j^2	(square)
6	$1/V_j$	(reciprocal)

The method for using SAM is similar to that for CORA. A main pro-

gram is written to read identifying names into AME, data into the storage array V and integers into the 1-dimensional arrays NR and NC specifying data selections by row and column. The first column of V specified by NC becomes the dependent Y and the following columns the independents X_1, X_2 ..., to a maximum of 9 i.e. to X_9. The selected variables are transformed by specifying integer values for the array LC, as above, in the same sequence as used for NC i.e. first value is for transformation for Y, second for X_1, and so on. The subroutine is then called by

```
CALL SAM (a, b)
```

Parameter a provides an option to obviate the need to specify rows (by integers in NR) when all the first N rows of data selected from V is to be used for the computations. Putting a = N indicates that all of the first N rows are to be used whereas putting a = 0 indicates that only the rows nominated in NR are to be used. Parameter b indicates the regression model to be used, using the ML integers given above. b = 0 produces a listing of all of the variable names in AME for reference purposes. The integer values read into NC and NR should always be concluded by the entry of 0, since SAM counts the number of integers up to this entry, to determine the number of data for the selected variables, before calling the regression subroutine REG. These requirements are covered in the example program REGR (microfiche).

The regression computations for variables selected, transformed, and arranged in one of the model forms by SAM, are carried out by the subroutines REG, INV and PRBF. REG provides the regression computations, INV inversions of symetric matrices and PRBF the probabilities of variance ratios (F - ratios). These subroutines may be used independently of SAM by writing a program to read dependent variable data into the array Y and independent variables into columns of X. These variables are specified in a common statement (see microfiche listing),

```
COMMON Y(200), X(200,20), BO, B(20)
```

which must appear in both the main program and the subroutine. BO and B(20) contain computed values for the b_0 and b_1, b_2 ... regression coefficients and may be used for further computations by the main program, for example for the computation of optimal fertilizer rates (Colwell, 1975) from the coefficients for a fertilizer - yield function computed by REG from fertilizer experiment data. The calling statement is

```
CALL REG (n, k)
```

where n is number of data in Y and columns of X, and k is the number of independent variables, or columns of X containing data. With the above common statement dimensions n and k have maximum values 200 and 20.

The subroutines SAM, REG, INV and PRBF are also listed on the accompanying microfiche film, together with an example main program REGR. A portion of the data bank read by REGR is also given, to illustrate the use of a simple main program, and example printed outputs produced by SAM and REG. The data bank values are followed by the data records containing integers for rows to be selected from V (NR entries), for the parameter ML indicating model, for columns of V (NC entries) and respective transformations (LC entries). Thus the concluding statements of the main program are

```
CALL SAM (0, 0)
```

to produce a listing of all variables in AME,

```
READ (LR, 190) GR, UP, N, (NR(I), I= 1,N)
```

to read a selection described by alphanumeric variables (GR, UP) and N values for row selections (NR), and finally a loop to read successive selections of models, variable selections and transformations,

```
READ (LR,210) ML, (NC(I), LC(I), I=1,10)
```

followed by the calling statement

```
CALL SAM (0, ML)
```

to produce the respective regression outputs. The example entries thus provide regressions and analyses of variances for the relationships

$$(Y\text{-}LN) = b_0 + b_1 \log(NO_3\text{-}M)$$

$$(Y\text{-}LN) = b_0 + b_1 (TAMM\text{-}FE) + b_2 (TAMM\text{-}FE)^2 + b_3 \log(NO_3\text{-}M)$$

$$(Y\text{-}LN) = b_0 + b_1 \log(TAMM\text{-}FE) + b_2 \log(NO_3\text{-}M)$$

$$(Y\text{-}LN) = b_0 + b_1 \log(TAMM\text{-}FE) + b_2 \log(NO_3\text{-}M) + b_3 \log(TAMM\text{-}FE).\log(NO_3\text{-}M)$$

$$(Y\text{-}LN) = b_0 + b_1 (TAMM\text{-}FE)^{\frac{1}{2}} + b_2 \log(NO_3\text{-}M)$$

$$(Y\text{-}LN) = b_0 + b_1 (TAMM\text{-}FE) + b_2 (NO_3\text{-}M) + b_3 (TAMM\text{-}FE)^2 + b_4 (NO_3\text{-}M)^2$$

representing a range of models for the relationship between yield response to N fertilizer (Y-LN), TAMM's acid oxalate soluble iron (TAMM-FE), and mean profile nitrate-nitrogen (NO_3-M), the bracketted names being those stored in AME. The analyses of variance show that the relationship is best represented by the second of these models. Such results usually suggest many other possible relationships which are easily tried by providing alternative data entries at the end of the data bank, as in this example. The subroutine SAM thus provides a very simple means for exploring regression relationships amongst extensive sets of data, such as those contained in the

National Soil Fertility Data Bank.

Microfiche

The accompanying microfiche contains program, data and output listings in the following sequence:

KO3	Main program CORS
MO3	Subroutine CORA
NO3	Subroutine CORREL
PO3	Function PRBF
BO4	Output of example program CORS, illustrating use of CORA
JO8	Data from NSFP data bank
C10	Main program REGR
E10	Subroutine SAM
I10	Subroutine REG
L10	Subroutine INV
O10	Function PRBF
O10	Output of example program REGR, illustrating use of SAM

References

Colwell, J.D. (1970). A statistical - chemical characterisation of four great soil groups in southern New South Wales based on orthogonal polynomials. Aust. J. Soil Res. *8* : 221-238.

Colwell, J.D. (1971). Effects of variations in soil composition on soil test values for phosphorus fertilizer requirements. Trans. Internat. Symposium on Soil Fertility Evaluation, I.S.S.S., New Delhi, Feb. 1971, pp. 327-336.

Colwell, J.D. (1977a). Soil testing and the efficient use of fertilizers. Reviews in Rural Science 3. University of New England, Armidale, N.S.W., Australia (in press).

Colwell, J.D. (1977b). Computations for studies of soil fertility and fertilizer requirements. CSIRO, Australia (in press).

Hallsworth, E.G. (1969). The measurement of soil fertility: the national soil fertility project. J. Aust. Inst. agric. Sci. *35* : 78-89.

Veldman, D.J. (1967). Fortran programming for the behavioural sciences, p. 131. Holt, Rinehart and Winston.

The use of a computer in a commercial soil analysis service

G.H. Price. Consolidated Fertilizers Ltd., Brisbane, Queensland

Introduction

In 1968 Austral-Pacific Fertilizers Ltd, a new fertilizer company based in Brisbane, began a 'Computerised Soil Testing Service'. Soils were sampled by field representatives and sent to the company laboratory in Brisbane for analysis. The results were printed by computer and sent back to the field man. Initially the results were printed by an IBM 1130 computer under a leasing arrangement with IBM, from 1969 on their own ICL 1901A computer. This computer was a 16K machine with two exchangeable disc drives, each of 8.2 million characters of storage. The card reader operated at 300 cards per minute and the printer at 600 lines per minute. The program for the soil testing service was written in FORTRAN and used about 14 000 words of memory.

Operation

Field sales staff collected representative soil samples from growers' paddocks. At the same time these representatives filled in a field information sheet for each sample (Fig. 1). The sample and corresponding information sheet were then sent to the laboratory where they were identified by date and a number. All information was punched onto cards (four cards per information sheet).
The initial intention was to develop a completely computerised soil testing service with the most modern facilities in Australia. Measurements for some nutrients were carried out on a routine basis for the first time in Australia. Analytical methods were automated and 15 analyses on each sample provided the basis for the service. These analyses were phosphorus (2 analyses, extracted with dilute H_2SO_4 and $NaHCO_3$ respectively), organic carbon, nitrate nitrogen, pH, conductivity, chloride, and water-soluble plus exchangeable sodium, potassium, calcium, magnesium, iron, copper, manganese and zinc.

After the soil sample had been analysed, the hand-recorded laboratory results were punched onto cards and read into the computer along with the field information. It matched the two sets of data according to laboratory date and number and then printed six copies of the field information and results for each sample (Fig. 2). A copy was sent to each of the following: farmer, field sales officer, company agronomist, local agent, accounts section and the central file in head office. This procedure commenced in mid-1968 and continued till December 1971.

TERMS AND CONDITIONS OF SOIL TEST

Austral-Pacific Fertilizers Limited ("the Company") makes no warranties whatsoever in connection with this soil test including its conduct, results, tabulation, subsequent and prior interpretation and any statements relating to prospective crop performance made by the Company, its employees, sub-contractors or agents and all terms, warranties and conditions whether express or implied, statutory or otherwise are excluded by this agreement. The Company shall not be liable for loss or damage of any description arising from the negligence or default of its employees, sub-contractors or agents or otherwise and no oral statements or representations made by the Company, its employees, sub-contractors or agents shall vary or revoke or supersede this clause in any way. PROVIDED that the general exclusion of all terms conditions warranties and liabilities by this clause shall in no way be limited by the matters specifically mentioned in this clause. In the event that this sheet is not signed by the customer no contract shall be deemed to have been entered into until delivery of an invoice and the soil testing result sheet.

SAMPLING PROCEDURE

The basis for a successful soil testing program is a good soil sample. Listed below are some points to guide you in making a good sample.

1. Be prepared with soil bags, information sheet, sampling tools and plastic buckets for mixing the samples.
2. Walk over the paddock and divide into uniform areas based on slope, colour, drainage, texture, crop growth or past cropping and fertilizer treatments. Sketch these different soil types on a farm map.
3. To sample an area, walk a "Z" pattern or a "Zig-Zag" pattern. Avoid unusual areas such as old or present fence lines, eroded areas, salt affected areas, swamps, low spots, road beds, dung or urine patches, or spots where piles of timber have been burnt. Problem areas may be sampled separately.
4. A composite soil sample is made by mixing 30-40 cores or plugs. To take a core or plug, use the recommended sampling tool and scrape away surface debris but do not remove top soil. When deep cores are taken, divide as recommended by our local agronomist. When all the cores or plugs are taken, mix well and fill the sample bag to the level indicated.
5. Give a number to each sample and record this number on the information sheet.
6. Fill in this sheet completely and enclose it in the rear pocket of the soil sample bag **so that the Farmer's Name and Sample No. can be seen.**
7. Store and despatch the samples as directed by your Austral-Pacific representative.

CLIENT'S AUTHORISATION — CUSTOMER MUST SIGN HERE

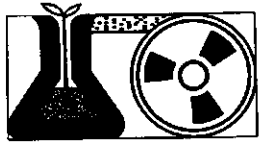
The soil test results are given subject to the terms and conditions stated above.

Signature.....Date / /

FC—011169

Fig. 1

COMPUTERISED SOIL TESTING SERVICE



SOIL TEST INFORMATION FIELD CROPS

FOLD HERE

TICK HERE IF DATA IS COMMON TO PREVIOUS SAMPLE

LABORATORY USE ONLY *day mth.*

LABORATORY DATE: 2 LABORATORY NUMBER 5 SUB-SOIL 9 10

9 INITIALS 15 NAME

35 ADDRESS

57 ADDRESS 72 STATE 76 POSTCODE

FARMER'S SAMPLE NUMBER 11 PADDOCK NUMBER 3 15 CUSTOMER NUMBER 23 DISTRICT CENTRE REP

CUSTOMER INPUT INFORMATION - TO BE COMPLETED ONLY FOR NEW CLIENTS

District Centre Territory CUSTOMER ACTIVITY LOCATION CODE

AGENT NUMBER CREDIT LIMIT

ACCOUNT NAME

Fold and pack with this side visible

LATITUDE 25 | LONGITUDE

TELEPHONE: 36 | TAKEN BY: 1. Company Representative 2. Farmer 3. Agent 4. Consultant 5. Company Contractor 6. Distributor

SAMPLING DATE: DAY MTH. 40

FALLOW RAINFALL, INCHES 50 | AVERAGE ANNUAL RAINFALL, INCHES

MONTHS OF FALLOW 52 | DEPTH OF MOISTURE, INCHES

DROUGHT 1. often 2. seldom 3. never 59 | FLOODING 1. often 2. seldom 3. never

IRRIGATION: 1. none 2. limited 3. plenty 61 | DRAINAGE: 1. poor 2. fair 3. good

SLOPE: 1. steep 2. irregular 3. gentle 4. level 83 | CONTOURED? 1. yes 2. no

COMMON SOIL NAME 85 | NORTHCOTE'S CLASSIFICATION

SOIL COLOUR 9 | SOIL TEXTURE 1. Sand 2. Sandy Loam 3. Silty Loam 4. Clay Loam 5. Clay

LOCAL CODING 15 | SOIL DEPTH INCHES

LAST STRAIGHT N APPLICATION MTH. YR 22 | LB./ACRE | FORM 1. Urea 2. Sulphate of Ammonia 3. Anhydrous Ammonia 4. Aqua Ammonia 5. Ammonium Nitrate 6. Calcium Ammonium Nitrate 7. Other

LAST OTHER FERTILIZER APPLICATION MTH. YR. 30 | LB/ACRE | FORM *

* Where possible use NPK Analyses, otherwise use abbreviations below.

OSP = Ordinary Superphosphate	APN = Ammonium Phosphate Nitrate
DSP = Double Superphosphate	RP = Rock Phosphates
TSP = Triple Superphosphate	AP = Ammoniated Phosphates
MAP = Monoammonium Phosphate	NP = Nitric Phosphates
DAP = Diammonium Phosphate	KN = Potassium Nitrate
APS = Ammonium Phosphate Sulphate	MP = Muriate of Potash
	SP = Sulphate of Potash

TOTAL P, LAST 5 YEARS, LB. 43 | YEARS CONTINUOUSLY CULTIVATED TO DATE

ORIGINAL VEGETATION 48

PAST CROPPING AND MANAGEMENT HISTORY — 3 years ago 54, 2 years ago, 1 year ago, Last six months: 1. Grass Pasture 2. Legume Pasture 3. Lucerne 4. Rice 5. Sugar 6. Winter Cereal 7. Summer Crop 8. Virgin 9. Short Fallow 10. Long Fallow 11. Small Crops 12. Other

STUBBLE/TRASH 64: 1. Nil 2. Burned or to be 3. Light Standing 4. Light turned in 5. Medium Standing 6. Medium turned in 7. Heavy Standing 8. Heavy turned in

WAS STUBBLE LEGUME BASED? 65: 1. Yes 2. No

WEEDS 66: 1. Nil 2. Moderate 3. Severe

PREVIOUS BEST YIELD 67 | 1. Lb./Acre 2. Ton/Acre 3. Bush./Acre 4. Other

ESTIMATED SELLING PRICE dollars cents 72 | UNITS 1. $ per lb. 2. $ per ton 3. $ per bushel 4. Other

COMPUTER PROGRAM NUMBER 76

PROPOSED CROP 3 | 0. Plant 1. 1st Ratoon 2. 2nd Ratoon 3. 3rd Ratoon 4. 4th Ratoon

VARIETY 20 | MTH. YR. PLANTING DATE

PLANTING RATE 32 | 1. Lb./Acre 2. Thousand Plants/Acre

ROW WIDTH, INCHES 36 | ACRES PLANTED

USED FOR 1. Grazing 2. Grain 3. Conservation 4. Other 41 | APPLICATION EQUIPMENT

FERTILIZER PLACEMENTS 51, 52: 1. Pre Plant 2. Banded with Seed 3. Banded away from Seed 4. Side Dress 5. Top Dress 6. Other

DESPATCH FROM CODE

DATE LAST PESTICIDE USED MTH. YR. 39 | PESTICIDE NAME

Economic analysis for an alternate crop (if required).

COMPUTER PROGRAM NUMBER 72 | dollars cents ESTIMATED SELLING PRICE

[illegible]
[illegible] & K.D. CANNON/BIMGARA
C/-G.CASSIDY/PO BOX744
BUNDABERG. QLD 4670

COMPANY REPRESENTATIVE - GREG CASSIDY (AGRONOMIST)
LABORATORY DATE - 100969
LABORATORY NUMBER - 1
DISTRICT - 2 111

0 SORGHUM R6

FARMERS SAMPLE NO. - 01 | PADDOCK NO. - SW1
TELEPHONE - 27945 | SAMPLING DATE - 8/ 9/69
AV. 9 MTHS RAINFALL - 28 INCHES | FALLOW RAINFALL - 18 INCHES
MONTHS OF FALLOW - 3 | DEPTH OF MOISTURE - 36 INCHES
PROVENT - SELDOM | FLOODING - NEVER
IRRIGATION - NONE | DRAINAGE - GOOD
SLOPE - GENTLE | CONTOURED - NO
SOIL - BY541 92X | YEARS CULTIVATED - 40
SOIL COLOUR - GREY | TEXTURE - SILTY LOAM
LOCAL CODING - GOOD STRUCTURE | DEPTH OF SOIL - 60 INCHES
LAST STRAIGHT N APPLICATION - | 11.68/ 56/UREA
LAST OTHER FERTILIZER APPLICATION - | 5.69/ 80/DIAMMONIUM PHOSPHATE
TOTAL P LAST FIVE YEARS - 60 LB. | ORIGINAL VEGETATION - BOX

THREE YEARS AGO - SUMMER CROP | TWO YEARS AGO - SUMMER CROP
LAST YEAR AGO - SUMMER CROP | LAST SIX MONTHS - SHORT FALLOW
STUBBLE/TRASH - HEAVY TURNED IN | NOT LEGUME BASED
WEEDS - NIL | PREVIOUS BEST YIELD - 65 BUSH./ACRE
ESTIMATED SELLING PRICE - | 0.00
PROPOSED CROP - SORGHUM R6
VARIETY - 612
PLANTING DATE - 11/69 | PLANTING RATE - 6 LB./ACRE
ROW WIDTH - 14 INCHES | ACRES PLANTED - 86
UTILIZATION - GRAIN | APPLICATION EQUIPMENT - COMBINE
FERTILIZER PLACEMENT - | PRE PLANT
LAST PESTICIDE USED - 6/68 | DDT

SOIL TEST RESULTS - SURFACE SOIL

PHOSPHORUS (PA)	69	PPM
PHOSPHORUS (PB)	13	PPM
ORGANIC CARBON (C)	2.3	%
NITRATE NITROGEN (N)	7	PPM
PH	6.2	
CONDUCTIVITY	0.06	MMHO
CHLORIDE (CL)	60	PPM
SODIUM (NA)	82	PPM
POTASSIUM (K)	110	PPM
CALCIUM (CA)	3260	PPM
MAGNESIUM (MG)	611	PPM
IRON (FE)	79	PPM
COPPER (CU)	3.7	PPM
MANGANESE (MN)	15	PPM
ZINC (ZN)	1.5	PPM

TO OBTAIN THE BEST RESULTS FROM THE ABOVE INFORMATION, AN AUSTRAL-PACIFIC FERTILIZER INVESTMENT PROGRAMME WILL BE DRAWN UP FOR YOU BY GREG CASSIDY
HE WILL BE CONTACTING YOU SHORTLY.

RECOMMENDATION

TEST STRIP 1
TEST STRIP 2
CONTROL STRIP

D. J. PARRY
CHIEF AGRICULTURAL CHEMIST

77

Fig. 2

In addition the computer carried out various calculations on each set of results. The field agronomists had devised correlation tables from experimental data and local knowledge for some crops on some soil types in their districts (see e.g. Fig. 3). These tables attempted to relate yield response to various soil test levels. This information was stored in the computer and was called on to interpret soil test results in the district, crop and soil type combination designated.

The calculations were presented as a 'fertilizer investment program' (see e.g. Fig. 4). The print-out gave the analytical results, the nutrient needs for a range of specified yield levels, fertilizer formulations based on urea, diammonium phosphate and potassium chloride, and the cost and return for the fertilizer recommendations at each yield level. Copies of these data were sent to the field sales officer and agronomist for the district to assist them with interpretations. The final recommendations were made in the field with the field sales officer discussing the program with the farmer. This procedure began in early 1970 and stopped in March 1971, when Austral-Pacific and ACF and Shirleys merged to form Consolidated Fertilizers Ltd.

Other ideas for improving the service were under consideration at that time but were not implemented. For example, it was intended that all instruments be fitted with paper tape punches to produce output for reading into a Nova Fairchild minicomputer located in the laboratory, which would carry out preliminary collating and batching prior to final processing on the 1901A. Thought had been given also to producing graphical output as a means of showing the grower which nutrients were most deficient. A district/crop/soil type standard output was to have been used as a base for evaluating the grower's results.

Problems

1. The company employed agricultural people with little computer background and computer programmers who had little or no biological or agricultural background.
2. The computer side of the service just grew as new ideas were thought up by the agriculturally oriented staff in the company. They knew they wanted to collate, store, print and recall the information but they were not sure how they wanted to use the data beyond this. Thus they could not indicate to the programmers definite future uses for the data. In turn, the programmers could not devise programs suited to ill-defined needs. As a consequence the programs were limited in their scope and often data could not be retrieved in the form ultimately decided upon.
3. Other problems concerned the interpretation of the results in the field. In many cases, there was doubt about the reliability of the nutrient response data on which the calculations for interpretation were made. In other cases, the N, P, and K requirements for a crop were calculated on the basis of a single application; this was quite impractical for crops which require split applications.

TABLE NO. - 101

CROP - WHEAT
SOIL TYPE - HEAVY
UNIT - BUS./AC
CENTRE - DARLING DOWNS

DATE - 13/08/70

LINE	DESCRIPTNS.	YIELD LIMITS 1	2	3	4	5	6	7	8	******** CODES 9	10	11	12
11	P(A) SOIL TEST	11.00	17.00	23.00	29.00	35.00	45.00	60.00	90.00	0	0	0	0.0
12	P FERTILIZER	0.00	3.00	6.00	10.00	15.00	20.00	25.00	30.00	0	0.	0	0.0
21	P(B) SOIL TEST	4.00	10.00	16.00	22.00	28.00	34.00	40.00	45.00	0	0	0	0.0
22	P FERTILIZER	0.00	3.00	6.00	10.00	15.00	20.00	25.00	30.00	0	0	0	0.0
31	STUBBLE FACTOR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
32		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
41	N SOIL TEST	4.00	9.00	13.00	18.00	22.00	31.00	36.00	41.00	5	0	0	0.0
42	N FERTILIZER	20.00	40.00	60.00	80.00	100.00	130.00	160.00	190.00	0	0	0	0.0
51	MOISTURE	12.00	15.00	18.00	24.00	30.00	40.00	50.00	60.00	0	0	0	0.0
52		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
61	PR SOIL TEST	2.50	3.00	3.50	4.00	4.50	5.00	5.50	6.00	0	0	0	0.0
62	SUBACTION	12.00	11.50	11.00	10.50	10.00	9.50	9.00	8.50	0	0	0	0.0
71		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
72		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
81	TOTALS SALTS	0.41	0.38	0.35	0.32	0.29	0.26	0.23	0.20	0	0	0	0.0
82	SUBACTION	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
91	TATAL SALTS	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0	0	0	0.0
92	SUBACTION	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
101	CL SOIL TEST	750.00	700.00	650.00	600.00	550.00	500.00	450.00	400.00	0	0	0	0.0
102	SUBACTION	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
111	CL SOIL TEST	1550.00	1500.00	1450.00	1400.00	1350.00	1300.00	1250.00	1200.00	0	0	0	0.0
112	SUBACTION	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
121	NA SOIL TEST	750.00	700.00	650.00	600.00	550.00	500.00	450.00	400.00	0	0	0	0.0
122		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
131	K SOIL TEST	80.00	100.00	120.00	140.00	160.00	180.00	200.00	220.00	0	0	0	0.0
132	K FERTILIZER	0.00	15.00	30.00	45.00	60.00	75.00	90.00	105.00	0	0	0	0.0
141	CA SOIL TEST	225.00	300.00	375.00	450.00	600.00	800.00	1000.00	1200.00	0	0	0	0.0
142	CA FERTILIZER	50.00	75.00	100.00	125.00	150.00	175.00	200.00	225.00	0	0	0	0.0
151	MG SOIL TEST	115.00	150.00	185.00	225.00	300.00	400.00	500.00	600.00	0	0	0	0.0
152	MG FERTILIZER	15.00	30.00	45.00	60.00	75.00	90.00	105.00	120.00	0	0	0	0.0
161	FE SOIL TEST	0.00	1.00	2.00	3.00	4.00	6.00	8.00	10.00	0	0	0	0.0
162	FE FERTILIZER	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
171	CU SOIL TEST	0.10	0.10	0.10	0.10	0.10	0.20	0.30	0.40	0	0	0	0.1
172	CU FERTILIZER	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
181	MN SOIL TEST	2.00	3.00	4.00	5.00	6.00	8.00	11.00	15.00	0	0	0	0.0
182	MN FERTILIZER	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.0
191	ZN SOIL TEST	0.50	0.60	0.90	1.20	1.50	2.00	2.50	3.00	0	0	0	0.0
192	ZN FERTILIZER	2.00	5.00	8.00	11.00	14.00	17.00	21.00	25.00	0	0	0	0.0
201	BUS./AC	10.00	20.00	30.00	40.00	50.00	60.00	70.00	80.00	0	0	0	0.0

Fig. 3

1 3 1
HARRY SMITH
P.O. BOX 210
DALBY 4405

1 1 11
JOCK HART
P.O. BOX 623
TOOWOOMBA 4350

AUSTRAL-PACIFIC FERTILIZERS LTD. - SOIL TEST INTERPRETATION

LABORATORY DATE - 100969 9142943/ 112 BORONUN
LABORATORY NUMBER - 1 MR. R.M. CANNON, LIGHT
FARMER'S SAMPLE NO. - 01 M.S. 263,
PADDOCK NUMBER - SW4 JONDARYN, QLD 4270

SOIL TEST RESULTS -	P-A	P-B	OC	NO3	PH	SALTS	CL	NA	K	CA	MG	FE	CU	MN	ZN	H2O
SURFACE -	69	13	2.5	7	6.2	0.06	60	820	110	3260	411	79	3.7	15	1.5	36
SUB -				2	6.5	0.06	80									

YIELD POTENTIAL -	N	P	K	CA	MG	FE	CU	MN	ZN	SALTS	PH	H2O	OLF
	34	37	106	160	160	160	160	160	100	0	160	80	160

YIELD - BUS./AC	NUTRIENT REQUIREMENT (LBS/ACRE) N	P	K	CA	MG	FE	CU	MN	ZN	GRADE %N	%P	%K	FORMULATION %UREA	%DAP	%MAP	%KCL	LB. PER ACRE	TON PER 86 ACRES
20.	0	0	0	0	0	0	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000
40.	4	0	0	0	0	0	0	0	0	46.0	0.0	0.0	100.0	0.0	0.0	0.0	8.7	0.334
60.	19	5	0	0	0	0	0	0	0	33.6	8.8	0.0	55.8	44.2	0.0	0.0	56.5	2.170
80.	44	10	0	0	0	0	0	0	0	34.9	7.9	0.0	60.3	39.7	0.0	0.0	126.1	4.841
100.	69	15	0	0	0	0	0	0	0	35.3	7.7	0.0	61.7	38.3	0.0	0.0	195.7	7.512
120.	99	25	14	0	0	0	0	0	2	31.0	7.8	4.4	52.1	39.1	0.0	8.8	319.3	12.259
140.	129	35	34	0	0	0	0	0	5	28.4	7.7	7.5	46.6	38.5	0.0	14.9	455.0	17.467
160.	159	45	64	0	0	0	0	0	8	26.0	7.4	10.5	42.2	36.8	0.0	21.0	610.6	23.443

TARGET YIELD BUS./AC	FERTILIZER INVESTMENT PER ACRE $	TOTAL RETURN PER ACRE @ $ 1.15 PER BUS. $	GROSS RETURN FROM FERTILIZER $	PROFIT FROM FERTILIZER $	INTEREST ON INVESTMENT %
20.	0.00	23.00	23.00	23.00	0 **
40.	0.32	46.00	46.00	45.68	14450
60.	2.31	69.00	69.00	66.69	2889
80.	5.09	92.00	92.00	86.91	1707
100.	7.87	115.00	115.00	107.13	1360
120.	12.68	138.00	138.00	125.32	988
140.	17.84	161.00	161.00	143.16	802
160.	23.59	184.00	184.00	160.41	679

PRICES USED ($ PER TON) - GIBSON ISLAND UREA = 81.44 DAP = 104.10 MAP = 96.65 KCL = 66.00

THIS IS A GUIDE ONLY. CERTAIN AGRONOMIC FACTORS WILL NEED TO BE CHECKED OUT WITH YOUR AGRONOMIST.
THIS SHEET MUST BE RETAINED WITHIN THE COMPANY.

Fig. 4

Finally, the models used for fertilizer economic calculations produced unrealistic recommendations in some cases.

Reasons for discontinuing computer processing

The company's computer was (and still is) used for many purposes besides the soil testing service. It carried out routine administrative functions such as accounting, payroll preparation, etc. It could only handle one job at a time.
For the soil testing service it was used virtually as a typewriter and provided very little benefit when used for this task alone. Sample numbers were low subsequent to the merger of the two companies and it was found that a typist could do the job at a lower cost than the computer. A job was considered uneconomical unless it ran for 30 minutes, yet it was calculated in 1971-72, when 3000 commercial samples were handled, that only 42 minutes of total printing time (in one run) was needed to prepare the result sheets for the whole year.

The cost of recording the laboratory information on punch cards, verifying, correcting, processing and transferring to disc or tape for filling for future retrieval would have been approximately $ 4000 (at 1972 values) for one year, providing the printing was done in one batch only. It would have required a minimum of 100 analysis sheets per day to warrant a daily print-out of results and even at this rate the printing time required was much less than the desired half hour. In short, it proved to be uneconomic to use the computer under the conditions which prevailed in 1972.

Future possibilities

Staff involved in the present soil analysis service are aware of the problems and pitfalls encountered in its early days. However, it is anticipated that the computer will be used again in the foreseeable future. The service has expanded recently, with sulphate sulphur, soil colour and texture assessed in addition to the 15 tests mentioned previously. The first step will be to define the ways in which the data will be used and then to design a system to achieve the goals set as efficiently as possible.
The data, consisting of 40 items of morphological and site information as well as the 16 chemical analysis, could be used in a number of areas of interest to the company and its customers.

In *research* the computer could be used to sort and group information into categories which can give, e.g. (a) trends in soil chemical properties and nutrient levels, in small districts or large regions, (b) frequency of occurence of particular soil types, soil chemical properties, crop sequences, fertilizer practices, etc., and (c) long-term trends in fertilizer use (nature, forms and rates of nutrients) in relation to crops and soil types in particular districts. The computer could be used to plot data on overlays of soil or other maps to show areas of high, medium and low levels of particular nutrients. In these types of activity, however, it must

be recognized that sampling for soil testing is biased and we should attempt to define the biases involved. For example, samples come from growers who have more than the average awareness of the benefits of improved technology.
Nevertheless, ready retrieval and manipulation of this data should enable the company to assess long-term fertilizer needs of particular crops, districts and soil types, so that its manufacturing, distribution and marketing groups can be given advance warning of the needs of particular districts or industries. Information on how many and which clients use soil analysis would be readily available. To be really useful the information would need to be updated frequently.

In the *marketing* of the company's products, computerisation of the soil analysis service should help to project a favourable image in the field. However, it should not be used for this purpose alone; the role of this soil analysis service is to help sell fertilizer where it is needed. The computer is a tool only and cannot replace the field adviser.
As in the past, the grower should receive his field information and analytical results on a sheet printed by the computer, to be interpreted by the company representative in the field discussion with him. They should decide together on a suitable fertilizer program for the particular crop.

The *design* of the new system would differ markedly form the old one as the company now possesses an ICL 1901T computer (64K) with two magnetic tape drives, four exchangeable disc drives, each of 60 million characters storage and faster input/output devices. It has an interactive time-sharing system and there are at present six (of a possible 48) remote terminals on-line to the computer. Although computing capacity is more than adequate to handle all the work done by the soil analysis service plus related research, the cost of using the computer for these purposes has not yet been assessed.

Conclusion

The computer can be a very useful tool to collate, group and correlate large amounts of agricultural data of considerable potential benefit to growers, government departments, research institutions and companies involved in the agricultural sector.
It is hoped that Consolidated Fertilizers Ltd will be able to contribute to Australia's fund of infromation on soils, soil fertility, and fertilizer needs and use. Future requirements for plant nutrients and soil amendments could conceivably be assessed with the help of data such as that held at present by the company if it could be made more readily accessible than is presently the case.

Use of a storage and retrieval system in soil fabric analysis

J.R. Sleeman. Division of Soils, CSIRO, Canberra, A.C.T.

Abstract

A storage and retrieval system for micromorphological descriptions of soils is briefly described and its use discussed. The system operates on data in coded form and has two distinct phases: the update phase in which the file of micromorphological descriptions is created, maintained, and added to or altered; and the search phase in which the file is examined to retrieve descriptions that satisfy given conditions. The output from the search phase consists of these slide designations together with their profile, horizon, and soil material classifications. The system has been used successfully to show the occurrence of specific fabric features within the various Great Soil Groups of Australia.

The setting up of the program and subsequent conversions have cost approximately $4000. Over a 5-year period 1370 descriptions have been added to the file for approximately $800; current rates are of the order of $1 for each description added. In this same period, 182 listings have been produced for approximately $100 i.e. approximately $0.60 per listing.

Introduction

The micropedology group of the Canberra Laboratories, CSIRO Division of Soils, has approximately 8 000 thin sections of soil materials, of which some 3 000 have been described in detail (see Stace et al. 1968). Each year this collection increases by 700 to 1 000 sections. The thin sections are sampled at particular depths (usually not more than 5 cm apart) to characterize and assess the origin of the profiles and the materials therein. The records of these profiles include profile and horizon classification, location of sampling site, site characteristics, macromorphology, physical and chemical analysis and micromorphorlogy.

The micromorphological descriptions are based on the system proposed by Brewer (1964). On average these descriptions consist of 60 words describing 15 fabric features out of a total of some 300 to 500 possible features. A typical description would be as follows:
Weak skel-vomasepic porphyroskelic fabric with channels, skew planes and vughs (C). Channel and plane ferri-argillans (C), with strong continuous or moderate orientation. Papules (R), with moderate to strong orientation, clustered in part. Normal red sesqui-

oxidic nodules (O). Irregular manganiferous nodules with sharp boundaries (O), clustered. Opal phytoliths (O). Black and brown opaque organic fragments (O). Pale brown anisotropic organic fragments with cell structures preserved (R).

These records contain information that could be useful in furthering our understanding of the conditions of formation and classification of fabric features, soil materials and profiles. To achieve this end one would require various listings of slides that contain particular fabric features or groups of features, the classifications of the soil materials and profiles that include them and the site characteristics of the sampled profiles. However, the compilation of even the simplest of these listings from the written records of the 3 000 thin sections currently described would probably take 2 to 3 weeks. As time passes the collection continues to grow and in effect the material becomes even less readily available. To rectify this situation a computer-based information storage and retrieval system was set up to handle the micromorphological descriptions and the various classifications of the soil material and profile within which the material occurs.

The storage and retrieval system

The SOLFAB system (Norris et al. 1971) operates on data in coded form and has two distinct phases: the update phase in which the file of micromorphological descriptions and classifications is created, maintained, and added to or altered, and the search phase in which the file is examined to retrieve descriptions that satisfy given conditions. The storage medium for SOLFAB is magnetic tape which holds the program, the codelist and the file of coded descriptions. The file is operated on in batch mode, instructions being entered by punched cards or remote terminal.

The codelist

This is the list of codes with the corresponding verbal description. A code consists of a set number (current permitted range 41-360) defining some micromorphological property, plus a symbol indicating one of a number of states of the property; any one such combination is called a characteristic.
For simplicity in handling, in any one description of a thin section (slide) a particular property can only be coded in one state, but as more than one state may exist in one slide a special symbol has been added: $ indicates that several states of the property exist together. A further special symbol, ?, indicates that although the property is known to exist no record of its state is available. Some lines of the codelist as they would be punched on computer cards are given below.

```
T       SESQUIOXIDE NODULES
178 A = RED SESQUIOXIDIC NODULES
178 B = YELLOW SESQUIOXIDIC NODULES
178 C = BLUE OR GREEN SESQUIOXIDIC NODULES
```

178 D = WHITE SESQUIOXIDIC NODULES
178 E = BLACK SESQUIOXIDIC NODULES
178 F = IWATOKA SESQUIOXIDIC NODULES
178 $ = SESQUIOXIDE NODULES COLOUR
178 ? = SESQUIOXIDE NODULES COLOUR
179 ∧ = 66 (referring to nature of boundary)
180 ∧ = 150 (referring to degree of adhesion)

The sets are grouped as necessary to describe classified groups, for example:

Glaebules	sets 172-273
Nodules	sets 172-242
Clay Mineral	sets 172-177
Sesquioxidic	sets 178-184
Manganiferous	sets 185-190

Group names and other comments which are solely for the guidance of the user are prefixed by T which produces spacing between successive groups in the printed codelist.
As the same property and states of that property are used to qualify different micromorphological classification groups, repunching of the list of states is avoided by the use of the symbol ∧ which indicates that, for example, set 179 has the same possible states as set 66; this concept has been called "equivalence". Thus the section of the codelist dealing with set 179 would be printed out by the computer as follows:

179 = 66
A = SHARP BOUNDARY
B = RATHER SHARP BOUNDARY
C = RATHER DIFFUSE BOUNDARY
D = DIFFUSE BOUNDARY
E = VERY DIFFUSE BOUNDARY

Coded descriptions

The coded descriptions of each thin section include four types of information: slide designation, date of coding, profile classifications, horizon and soil material from which the thin section was taken, and the description of the micromorphology in terms of the codelist.
The coded description for the thin section described in the Introduction is given below. The slide designation is C1/4 and the material was sampled at a depth of 34 cm in the B_2 horizon of a Red-brown Earth profile (Stace et al. 1968), Db 1.33 (Northcote 1971) or haplustalf (Soil Survey Staff 1960, 1967); the micromorphological description was coded on the 19th November 1969.

C 1/ 41 41 19/11/6920DB1.33 7.42 7 45 B 46G 48A 51? 55?
58B 69A 72$ 73$178A
C 1/ 42179A182A185A188B189B263A266B287B311B312A320$321E

The update phase

In this phase the codelist can be updated, and the file of coded descriptions can be added to, altered, deleted or printed. Six different forms of printed output may be produced depending on the control cards used.
1. A summary of the control cards chosen and a count of the number of descriptions given. A list of designations for these descriptions altered, added to, or deleted from the file together with the total number of descriptions in the new file.
2. A newly defined codelist.
3. A listing of the data following any control card.
4. A complete description, including both the code and related verbal terminology, for all additions free of errors.
5. A complete description of all altered micromorphological descriptions.
6. A complete description of all micromorphological descriptions required to be printed.

The search phase

In this phase information is extracted from the file by a search based on a criterion which may contain any logical combination of the following conditions:
1. that a particular characteristic exists. This is specified by using the code for that characteristic.
2. that any characteristic from a given set exists. This is specified by the set number followed by the special symbol.*
3. that a particular classification is stored in the description. The classification is specified by using a mnemonic representation for the classification, followed by the class code in brackets. The mnemonics are: HANDCLAS (handbook classification; Stace et al. 1968), NORTHKEY (principle profile forms; Northcote 1971), USDA7 (USDA soil classification; Soil Survey Staff 1960, 1967), HORGEN (genetic horizon; Soil Survey Staff 1951, 1962), HORDIAG (diagnostic horizon; Soil Survey Staff 1960, 1967) and SOMA (soil material classification, proposed by Brewer 1964).
4. that coding was carried out on, before, or after a given date. This is specified by one of the following mnemonics: DATEE, DATEB, DATEA. The date follows the mnemonic and is enclosed in brackets, e.g. DATEA (1/10/69) would select all those coded after 1/10/69.

The conditions may be linked to any degree of legical complexity using the logical connectors $\wedge$ and $\vee$, representing AND and OR. The symbol $\neg$, representing NOT, may be used to negate a condition. Brackets may be used to logically group the conditions.
For example if the requirement was a list of all slides from Red-brown Earth profiles including carbonate nodules but without sesquioxidic nodules the criterion would be:

HANDCLAS (20) $\wedge$ 197A $\wedge$ $\neg$ 178 *

The printed output from the search phase includes the searching

criteria, the date of search, and the slide designations satisfying the criteria together with their profile, horizon and soil material classifications.
For example:

FILE SEARCH - 01 / 09 / 70

CONDITION FOR SEARCH:

HANDCLAS (20) ∧ 197A ∧ ¬ 178 *

SLIDE NO.	HANDCLAS	NORTHKEY	USDA7	HORGEN	HORDIAG	SOMA
C 48/7	20	Dr 2.33		7		

Error checking facilities are incorporated into both phases (e.g. cards out of sequence, use of illegal characters or symbols, incorrect card format, etc.) and where applicable error messages are included in the printed output.

Using the system

Most hypotheses concerning the origin of fabric features and their relationship to soil behaviour have been derived from studies on a limited range of soils or soil materials. Such hypotheses can be considerably enhanced and extended by checking against a wider range of materials. This is facilitated by the use of a system such as SOLFAB which can recall relevant material for further consideration or investigation.

A current project in the micropedology group is the preparation of an atlas of typical fabric features as observed in thin sections of Australian soils. Photomicrographs of these features are to be accompanied by descriptions and comments on nomenclature, specific and general occurrence in relation to profile, horizon and site characteristics, and genesis. It would be very tedious to check written records for the location of all the features observed in the 3 000 thin sections described over the past seven years, let alone the relationship of these features to particular kinds of horizons, profiles or site characteristics. SOLFAB has made most of this information readily available for fabric features recorded in the descriptions of 1370 thin sections, representing 32 Great Soil Groups, currently held in the file. To date listings haven been prepared for 182 fabric features. Thse listings indicate the presence of a feature in the file, give a range of thin sections for the study of a particular feature, and reveal the relations, if any, with particular kinds of horizons or profiles. Personal inspection is still required to check correlations with site characteristics not included in the system. Some examples of information obtained through the use of the system are given below.
1. Manganiferous nodules are found to have a wide-spread occurrence, occurring as they do in 513 thin sections representing 20 Great Soil Groups. Most of these nodules are irregular with sharp boundaries and commonly occur throughout the profile; irregular diffuse

nodules are confined to solonetzic soils.
2. Glaebular halos are of restricted occurrence, occurring in only 42 thin sections from six Great Soil Groups of mildly to strongly leached soils that are generally sodic and may be saline.
3. Laminae have been recorded in 295 thin sections from 17 Great Soil Groups mainly formed on alluvial parent materials, giving support to the postulated origin as sedimentary features formed during the deposition of the transported parent material. The laminae not associated with alluvial parent materials are clay laminae that probably in fact are embedded argillans.

In relation to soil behaviour, materials with dominantly planar voids can be readily selected for testing and checking of the postulate of Lafeber and Willoughby (1967) that physical anistropies (in particular planar voids) significatly affect the behaviour of a soil under load.
The particular system described is not restricted to handling soil micromorphological data. Any data that can be fitted to the general form adopted here could be stored and accessed by this system. Information about diagnostic horizons of the USDA Soil Taxonomy (Soil Survey Staff 1975) could be stored; it would be easy to check, for example, whether any property is associated with a particular horizon. Similarly it might be used to store information about soil horizons as defined by Fitzpatrick (1967).

Costs

An approximate costing of SOLFAB from its initiation in 1970 until the end of 1975 is given below.

		$
PROGRAM	Initial setup (on CDC3600)	3 000
	Conversion to CDC Cyber 76	1 000
File	Currently holding 1370 descriptions	800
	(1970, $180 per 300 descriptions)	
	(1975, $250-300 " ")	
SEARCH	To date 182 listings made	100
	(1975, $40 per 60 listings)	
STORAGE OF MAGNETIC TAPES SINCE 1970		300
	($5 per month for 2 tapes)	
TOTAL COST 1970-75		5 200

It can be seen that the total cost represents a cost of $28 for each of the 182 listings made so far; however, each additional listing at this stage would cost less than $1. Looking at it in another way, the total cost is equivalent to 17 weeks (at $300 per week) of personal inspection of the written records. Whilst this amount of time may be sufficient to produce 182 listings, the system is capable of preparing many more listings at a cost of less than $1 each.
Thus as time goes on, with increasing size of the SOLFAB file and increasing use, the advantage of using a computer-based file rather than a manual one will increase.

References

Brewer, R. (1964). Fabric and Mineral Analysis of Soils. John Wiley and Sons, Inc., New York

Fitzpatrick, E.A. (1967). Soil nomenclature and classification. Geoderma 1:91-105.

Lafeber, D. and D. Willoughby (1967). Geometrical anisotropy and triaxial failure in soils. Proc. 3rd Asian Reg. Conf. Soil Mech. and Found. Eng., pp. 186-192.

Norris, J.M., S.W. Cumpston and J.R. Sleeman (1971). A storage and retrieval system for micromorphological descriptions of soils. CSIRO Aust. Div. Soils Tech. Pap. No. 10.

Northcote, K.H. (1971). A factual key for the recognition of Australian soils. pp. 123. Rellim Tech. Publs, Glenside, S.Aust.

Soil Survey Staff (1951). Soil Survey Manual. Handbook No. 18. U.S. Dept. Agr., Washington, D.C.

Soil Survey Staff (1960). Soil Classification: a Comprehensive System, 7th Approximation. U.S. Dept. Agr., Washington, D.C.

Soil Survey Staff (1962). Identification and Nomenclature of Soil Horizons. Suppl. to Handb. No. 18. U.S. Dept. Agr., Washington, D.C.

Soil Survey Staff (1967). Supplement to Soil Classification System (7th Approximation). U.S. Dept. Agr., Washington, D.C.

Soil Survey Staff (1975). Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys. Soil Conserv. Serv. U.S. Dept. Agr., Handb. No. 436, 754 pp. U.S. Govt. Printing Office, Washington, D.C.

Stace, H.C.T., G.D. Hubble, R. Brewer, K.H. Northcote, J.R. Sleeman, M.J. Mulcahy and E.G. Hallsworth (1968). A Handbook of Australian Soils. Rellim Tech. Publs, Glenside, S.Aust.

Comparative costs of data handling in land resource surveys

N.M. Dawson, Department of Primary Industries, Brisbane, Queensland
A.W. Moore, Division of Soils, CSIRO, Brisbane, Queensland

Introduction

It has become fashionable of late to disparage the use of computers as high-powered calculators or super-clerks. Nevertheless, the transliteration of manual data handling systems to electronic data processing has been one of the few really successful areas of computing up to the present time, in the sense that almost invariably electronic data processing eventually proves to be cheaper. It has been suggested that one of the myths of the computer industry is that 'it is cheaper to automate than innovate' (Sibley 1974). Our experience in using computer assistance for certain aspects of land resource survey suggests that it is usually cheaper to automate and that to some extent automation forces one to innovate.

In this paper we do not intend discussing the desirability or otherwise of carrying out land resource surveys, the philosophy behind such activities or operational procedures in the field. Rather we have taken an information system already in use and tried to compare the costs of certain data handling aspects when carried out by hand and by computer.

Costs of land resource surveys are high, with the major part of these costs being salaries. For example, salaries accounted for $125 000 of the total cost of $150 000 for the Western Arid Region Land Use Study (WARLUS) of the Queensland Department of Primary Industries (Anon. 1974). These figures indicate that savings in manpower would be an important factor in reducing costs in future surveys.
In most land resource surveys large amounts of data are collected at sample sites by experienced personnel. Much of this data is collected on the basis that costs and project efficiency will not allow the sites visited during the data collection stage to be revisited. Under these conditions survey teams aim at collecting as much relevant data at each site as is possible within a defined time schedule. In fact, experience has shown that invariably some of this data is not used in the formal report but proves invaluable to other users once they are aware of its existence.

Prior to 1970 most organizations involved in survey work in Queensland used both field books and site description sheets to record soil, vegetation and other land data. Information from these sources was generally processed manually, a job that was time-consuming and in many ways boring. The system currently used by

the Queensland Department of Primary Industries arose because of a desire to improve the efficiency of collection and retrieval of such land resource data.
At that time one of us (A.W.M.) had had some experience of using a generalized database management system (GDBMS) for storage and retrieval of data from past CSIRO soil surveys (Moore et al. 1974). It was felt that the initiation of WARLUS provided an opportunity to compare costs of using a GDBMS for that survey with costs of manual handling of data from a similar survey carried out in the past, viz. the Miles Technical Guide Study (MTGS) (Dawson 1972).

Comparisons were made as far as possible at five stages in the processing of data, viz. (a) planning and establishment, (b) data collection, (c) data input, (d) storage, and (e) retrieval and display. Because some common basis of comparison was necessary we have used 1972 Australian dollars in this presentation.

Database management

Most of the database management mentioned in this paper was done using INFOL, a generalized file management system written in COMPASS for the CDC 3600 computer. The facilities provided by 3600 INFOL are those frequently required by many people wishing to manipulate data, e.g. retrieval using sophisticated logic, sorting, file inversion, automatic or user-controlled format for output as hard-copy or on other media. It is a conceptually sound GDBMS, but falls short of perfection in implementation. Problems were in large measure due to lack of modularity in the system - debugging in one part usually resulted in rebugging in another.

With the advent of the CSIRO Cyber 76 computer a new version of INFOL written in FORTRAN has been made available by the Division of Computing Research. While this is a more reliable system, it is far less comprehensive than the original INFOL and its usefulness will remain limited unless further resources are allocated to its development.

There are trade-offs in the use of a GDBMS, compared with a program written to carry out a specific job. Where a system has a large number of users, which is more likely with a GDBMS, its development costs are amortized over all of them and thus become virtually negligible as far as a particular database is concerned.
For example, the cost of setting up the WARLUS file (excluding planning, which accounted for about 1/4 of the cost) was approximately $900 (see next section) using INFOL, which involves a non-procedural language. This can be compared with the relatively less complex soil micromorphology file system of Morris et al. (1971) written in FORTRAN, a procedural language, which cost at least $3000 to set up (Sleeman priv. comm.). The use of a non-procedural language also means that files can be set up by staff with limited programming skills.
On the other hand, the person who elects to use a GDBMS (assuming one is available) usually relinquishes a large amount of control

over the system. Further, the system may not provide certain required facilities so that interfacing with external programs becomes necessary.
Because of the unreliability of INFOL the Queensland Department of Primary Industries has over the last two to three years changed over to writing its own specific programs (Cormack priv. comm.). A generalized translation program to translate the fixed-field format was produced (Walker et al. 1973).

Planning and establishment

In the procedures directly involved in an information system, there are two major activities: (1) the establishment of data collection methods, and (2) the establishment and testing of a computer database system.
The major costs involved with the information system for WARLUS were associated with these two activities. The field data to be collected are basically determined by the objectives of the project and well-tried convention. The data collection form used in MTGS had been previously used in other surveys and was based on the CSIRO Soils Division standard recording sheet. However, there were differences in the way data were recorded by different officers in both the Department of Primary Industries and the CSIRO Division of Soils. These sheets concentrated mainly on soil characteristics whilst WARLUS placed added emphasis on other land features (see Table 1).

Two main constraints became evident when the number of attributes to be recorded were considered. These were the time which could be allotted to description in the field and the size of an entry (logical record) in the file.
Many hours were spent planning the classification of attributes to be recorded in WARLUS. Much of the time was spent ensuring that descriptions used were compatible with other surveys. In other cases currently used classifications of certain attributes were considered unsuitable as a result of previous experience or due to the fact that they were qualitative rather than quantitative. In the latter case, these were quantified where possible. Attribute classes were coded as it was decided that time could be saved by recording codes rather than longhand descriptions. Data description sheets and field recording sheets were then prepared, tested and printed. The initial field sheets, whilst being suited to field use, proved unsuitable for punch card operators so they were redesigned and the resultant field sheets (see Figs. 1 and 2) proved to be a satisfactory compromise for both field recording and card punching.

An INFOL file structure was established and data from a number of the MTGS sites were used to test the system. Eighteen man-days were involved in the over-all planning and establishment of this system. If a value of $50 per man-day is placed on these activities then the labour cost of establishing the system was $900. To this must be added the cost of computing ($200) giving a total cost of $1100.

LAND USE SURVEY RECORDING SHEET
DIVISION OF LAND UTILISATION
QUEENSLAND DEPARTMENT OF PRIMARY INDUSTRIES

SURVEY NAME WARLUS B

SITE NUMBER 8

LONGITUDE	LATITUDE	R502 SHEET	AIRPHOTO REFERENCE	GRID REF.	SURVEYOR	DATE	ANN. RAIN	L.F.	POS.	SLOPE	M.A.	ASP.	REL. RELIEF	ELEV.	SURVEY	SITE	CARD
1 2 3 4 5 6 7	8 9 10 11 12 13 14 15	16 17 18 19 20 21	22 23 24 25 26 27 28 29 30 31 32 33 34	35 36 37 38 39 40	41 42 43 44 45	46 47 48 49 50 51 52	53 54 55	56 57	58 59 60	61	62 63	64 65 66 67	68 69 70 71	72 73 74 75	76 77 78	79 80
1 4 5 0 5	E 2 5 3 1	S C 5 5 0 5	6 9 0 5 0 2 4 3 1	8 9 1 6 7 3	0 0 8 2 8	M I L J R	7 2 0	5 2 5		1 7	0		3		B Q	W A	8	1

GEOL. REF.	P.M.	GEOM.	WIND SEROS. WAT.	STONE BARE	ROCK	DRAIN.	RUNOFF	S. DEPTH	S. DEPTH	SURFACE CHAR.	BLEACH	PROFILE TYPE	LAYER NO.	SOIL SAMPLING DEPTHS	SOIL TEXTURE	MUNSELL COLOUR	SURVEY	SITE	CARD
1 2 3 4 5 6	7 8 9 10 11 12 13 14 15	16 17	18 19	20 21 22	23 24	25 26	27 28	29 30 31 32 33 34 35 36 37 38 39 40 41	42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71	72 73 74 75	76 77 78	79 80
Q A	1 1 2 L	F	0 3	0 4	C	3 2		5 F	C K A	0 5				1 0 20 30	6 0 1 2 0	1 8 2 1 2 1 2 1 2 0	4 0 4 4	4 0 3 4	4 0 3 4	9 0 4 4			2

MUN. COL.	SOIL COLOUR	MOT. TYPE	DEGREE OF MOTTLING	STRUCTURE	CONS-MOIST	CONS-DRY	FIELD pH	CONCRETIONS	SURVEY	SITE	CARD
1 2 3 4 5	6 7 8 9 10 11 12 13 14 15	16 17	18 19 20	21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40	41 42 43 44 45 46	47 48 49 50 51	52 53 54 55 56 57 58 59 60 61	62 63 64 65 66 67 68 69 70 71	72 73 74 75	76 77 78	79 80
4 0 4 4	1 7 1 7 1 7 1 7 1 7	[redacted]		3 2 B L 3 2 B L 3 2 B L 3	B L 3	B L	1 1 2 3 2	3 6 5 6 0 6 0 7 5 6 0			3

LAND SYSTEM	L.U.	FACET	G.S.C.	P.P.F.	VEGETATION SUMMARY	G.C.	EROS.	REGR.	FLOOD	SCALD	SAL.	MY. MOT	PERT.	PMY. SOIL	STONE	TOPO.	DRAIN.	MICRO.	DEPTH	RABBIT	TERMITE	POLISON BAIT	PHOTOGRAPH	SURVEY	SITE	CARD
1 2 3 4 5 6 7 8 9	10 11 12	13	14 15 16	17 18 19 20	21 22 23 24 25 26 27 28 29 30 31 32 33 34 35	36 37	38 39	40	41	42 43	44 45	46 47	48 49	50 51	52 53	54 55	56 57	58 59	60 61	62 63	64 65	66 67	68 69 70 71	72 73 74 75	76 77	78 79	80
1 4 4	U	6	5 3	4 A	C H A R E R M I T S W L M	1 1	2	1		0 0	1	0	1	0	2	0	0	0	0						1	0	

SOIL PROFILE

0 CMS.
10
20
30
60
120

NOTES

2hr auger of silty loamy clay platy / granular structure

Branched channels with bigalow / gidgee / sandalwood fringing bare open woodland on grey and brown clays with thin surface crusts

Fig. 1

VEGETATION CHARACTERISTICS SHEET

DIVISION OF LAND UTILISATION
QUEENSLAND DEPARTMENT OF PRIMARY INDUSTRIES

SURVEY NAME WARLUS B

SITE NO. 88

DIST.	D.	A.P.O.	P.GR.	A.GR.	P.F.	A.F.	T.F.	L.	NO. S	SPECIES 1–5	HT.	DIAM.	% COVER	DISTRIB.	FORM	RANK	SURVEY	SITE	CARD
	1	2	7	1	1	0	1	5		ACAARACCAMEUMICGEPARERMIT	66744	658	543R2R	CCCCC	LTLTLTHSLS	12453	BQWA	8	8

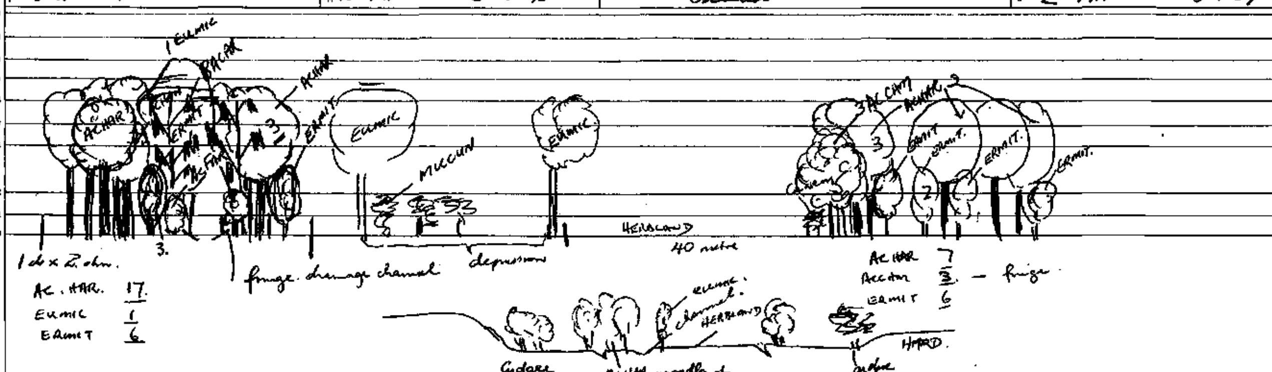
REGENERATION	EATEN	OTHER SPECIES	T/A	SURVEY	SITE	CARD
1RRRRRFRFRN	NUNN	CALØRACFARBACARSAKALBØDIFATMUEBACALBATRIAPAMMSØESUSPACT	A	BQWA	8	9

SPECIES PRESENT: CANEM CYBIF MASPI PAJUB JUNCUS BAVEN MUCUN – depression MADRU
CHACI SIDA POSPA ALNOD BABIL DARAD Opuntia ERMEM LEDIG

VEG. SUM. P.F.G. %: ACHAR S WLM

		CONDITION	
T 8-12m : 10-20 B 80-100	M 2-4m 5-10%	L Isolated	G <.7m 15-25%

PROFILE M >30, 30, 20, 15, 10, 7, 5, 2, 1, 0



NOTES: 1 ch x 2 chn.
AC. HAR. 17
EUMIC 1
ERMIT 6

PHOTO C | B/W | HERB. SPEC.

Fig. 2

It is worth noting that many of the costs associated with this project were innovative costs and have not been repeated for succeeding surveys which have used the same input format. In the case of MTGS the establishment costs were minimal as recording sheets and attribute descriptions were already available, these having been developed for previous studies.

Data collection

In WARLUS land characteristics were coded directly onto the newly-designed field sheets. In MTGS data was recorded in longhand on conventional field sheets. Initially there was resistance to the use of codes by field workers. However, once they became familiar with the codes and the method of operation was accepted efficiency improved rapidly. Whilst the amount of data collected per record was greater in WARLUS than in MTGS (see Table 1) it was found that the actual time involved in augering the soil hole and describing the soils and other land features remained approximately the same. Time spent at each sample site ranged from 20 to 60 minutes in both surveys, with the average time being approximately 30 minutes. The scope for increased efficiency was much the same in both surveys as the time taken at each site is largely determined by the time taken to auger the soil profile and collect plant specimens. In MTGS the field team comprised one of the authors (N.M.D.) and a technical assistant whereas WARLUS involved a botanist as well. Excluding the vegetation component of the two field records it can be seen (Table 1) that information about 88 attributes was recorded in WARLUS as against 54 in MTGS. While this improved efficiency of collection was not directly dependent on the introduction of computer processing, the latter did lead to a reexamination of what data to collect and to certain changes consequent on this. Details of the set of data collected in WARLUS have been listed by Dawson (1975).

There were both advantages and disadvantages in the use of coded data. The advantages were that coding in the field was quicker, more precise and involved discipline in classification. It was not possible to be loose in a description, nor was it possible to avoid description. This meant a positive decision or a compromise had to be made in the field, not later when recall was poor. As an example wind erosion at a given site may range from class 0 to 2. In WARLUS only one class could be recorded on the field sheet so the surveyor had to decide which class was most representative of a particular site. This in fact was also one of the disadvantages as conditions can show a wide range. In these situations the range was included as additional data at the bottom of the sheet but did not enter the computer record.
In the field it was found that this systematic approach was more efficient (in terms of number of items of information collected per unit of time) than the older approach even though actual costs per unit of time were much the same.

Table 1. Data collected by the Queensland Department of Primary Industries for two land resource surveys

Type of data	Miles Technical Guide Study (Dawson 1972)		Western Arid Region Land Use Study (Anon. 1974)	
	No. of attributes	No. of items/ subitems recorded	No. of attributes	No. of items/ subitems recorded
Land data	9	9	18	18
Soil data - general	3	3	12	12
Soil data - morphological	11	55	12	60
Soil data - analytical	21	105	21	105
Vegetation data	12	12	21	69
Interpretive data	6	6	17	17
Total	66	194	109	293

Data input

Data collected on WARLUS field sheets were punched directly onto cards, up to 10 per site. The formats used for these cards were the same as those on the recording sheets. The cost of this operation was 64 cents per site if the two vegetation cards are excluded (this makes the comparison with MTGS more reasonable). These cards were then listed at a cost of 1 cent per site for computer time. In the case of MTGS, field sheets were corrected and data completed by additions in the office. These sheets were then typed, the costs being similar to those for punching computer cards. Costs (i.e. time involved) for visually verifying both sets of data were much the same.
Both sets of records were corrected manually. It was clear at this stage that less mistakes were made in the transcription of the coded data to the punch cards (WARLUS) than in the typing of record sheets (MTGS). On other words, the field man's writing was bad, whilst his printing was reasonable.

The total cost of data input (listing, repunching, reformatting two cards) for the WARLUS file (225 records) was $80. As with collection it seems that costs were much the same for the two surveys but efficiency and accuracy were higher in WARLUS.

Storage

Costs of computer storage depend on the storage medium and the charges it attracts at any particular time. During the period the

WARLUS file was being used via INFOL the cost of entry was $14, $28 or $38 depending on whether CSIRO, State Department, or outside rates for computer time were used in accounting. However, all users were charged a flat rate for magnetic tape storage. For a primary tape + back-up tape the above input costs were equivalent to the cost of 4, 12, and 16 months' storage respectively early in the period, and 3, 6 and 9 months at a later date when tape rental increased. Thus the most economical method of storage depended on size of file, the somewhat arbitrary and varying charges for services provided by the Division of Computing Research, and the frequency of use of the file. Under some circumstances storage on punched cards was most economical but at all times storage on disc was prohibitive. Punched cards or manual records (e.g. those of MTGS) cost virtually nothing to store.

This highlights the fact that choice of a particular way of doing something is often not a technical decision but simply a matter of money. Further, when a user is locked into a computing facility such as that of the CSIRO Division of Computing Research in which charges vary arbitrarily (from the user's point of view), the economic viability of a particular information system can vary frequently and over very short time intervals.

Retrieval and display

The main objective of this experiment in database management was to increase the efficiency of data retrieval. It was hoped that this would enable more data analysis and interpretation to be carried out and result in considerable savings in costs and time, thus releasing more resources for idea generation.

From the WARLUS file coded data could be retrieved in either coded or decoded form. In most cases code was used as it was cheaper to display and readily understood by users after relatively short experience with it. Descriptions of different soil and vegetation groups were produced by sorting and simple print-outs of data. Costs of computer vs manual operations were compared for retrieval and sorting. Only the WARLUS sites were used, thus allowing operations on the same set of sites to be compared by the two methods. However, this comparison favoured manual interpretation because the WARLUS field sheets were better structured and more uniformly and consistentsly filled in than the MTGS sheets.

As mentioned previously, the performance of INFOL during the period of data interpretation was far from perfect and consequently only limited use of it could be made. This subsequently forced the users to interpret much of the data using other programs (written in FORTRAN) which proved to be cheaper, although not providing the same range of options. This did not, however, negate the cost saving (see below) but meant that optimal use of the data retrieval program was not possible because of operational problems.

In the simplest type of operation data was retrieved on a single

Table 2

6 JUN 1972

SITE NO.	SOIL SERIES GROUP	GREAT SOIL GROUP	PPF	SOIL TEXTURE	FIELD PH	SOIL SURF	COLOUR	VEGETATION SUMMARY
59		RED EARTH	GN2.11	15 15 19 19 19	5. 5. 5.5 6. 6.	H	30 30 30 30 30	ACANE ERDIE GOST
61		RED EARTH	GN2.11	15 19 19 19 19	5.5 5.5 5.5 6.5 6.5	H	30 30 30 30 30	ACANE ERLON OSH
43		RED EARTH	GN2.11	15 19 20	5.5 5. 5.	H	30 30 30	ACANE - OSH
134		RED EARTH	GN2.12	11 11 15	6.5 6.5 6.5	H	30 30 30	ACSPA - OSL
128		RED EARTH	GN2.12	11 14 19	7. 7. 7.5	KH	30 30 30	ACANE CASPP FOSH
25		RED EARTH	GN2.12	11 15 15 15	5. 5. 5.5 5.5	K	30 30 30 30	ERNUC EUTER WOHT
119		RED EARTH	GN2.12	11 15 15	6. 5.5 5.5	H	30 30 30	ACSPA - OSH
89		RED EARTH	GN2.12	11 15	6. 7.5	K	30 30	ACANE CADES SOST
107		RED EARTH	GN2.12	11 15	6.5 7.	H	30 30	BASPP ACANE SOHH
5		RED EARTH	GN2.12	14 14 14 14 17	5.5 6. 6. 6.5 6.5	LK	30 30 30 30 30	ACANE - GOST
101		RED EARTH	GN2.12	14 14 14 17 14	5.5 5.5 5.5 7. 7.5	KH	30 30 30 30 30	ACANE - FOSH
98		RED EARTH	GN2.12	14 14 14	5.5 5.5 7.	H	30 30 30	ACANE EVTER GOSH
31		RED EARTH	GN2.12	14 14 17 20	5.5 5.5 6. 6.5	KH	30 30 30 30	ACANE - GWLL
39		RED EARTH	GN2.12	14 15 15 17	6.5 6.5 6.5 6.5	KH	30 30 30 30	ARCON ACANE SOHT
40		RED EARTH	GN2.12	14 15 19 19 19	6.5 6.5 6.5 6.5 7.	KH	26 30 30 30 30	ACANE ERGIL SSHT
24		RED EARTH	GN2.12	14 15 19 19	6.5 6. 6. 7.	KH	30 30 30 30	ACANE ERERI OST
8		RED EARTH	GN2.12	14 19 19 19	5.5 5.5 6. 6.5	KH	30 26 26 26	ACANE ERGIL GOSH
129		RED EARTH	GN2.12	15 14 17 19 19	7. 7. 8. 7.5 7.5	LH	30 30 30 30 30	ACANE CASPP SOSH
114		RED EARTH	GN2.12	15 15 15 19	6. 6. 6. 7.	CH	23 23 23 23	ACANE EUTER GOST
115		RED EARTH	GN2.12	15 15 15 19	6. 6. 6. 7.	CH	30 30 30 30	ACANE DISER GOST
143		RED EARTH	GN2.12	15 15 15 19	6.5 7. 7. 8.	H	30 30 30 30	- - NOV
1		RED EARTH	GN2.12	15 15 19 19 19	6.5 6.5 6.5 7. 7.5	KH	23 26 30 30 24	ACANE ERGIL SSHT
2		RED EARTH	GN2.12	15 15 19 19 19	5.5 5.5 6. 6.5 7.	K	30 26 26 26 30	ACANE - GSHT
82		RED EARTH	GN2.12	15 15 19 19 19	6. 5.5 6. 6. 7.	H	30 30 30 30 30	ACANE ARSPP GOSC
125		RED EARTH	GN2.12	15 15 19 19 19	6.5 6.5 7. 7.5 8.	KH	30 30 30 30 30	ACANE EUTER GOSH
111		RED EARTH	GN2.12	15 15 19 19	6. 6. 6. 8.	CH	30 30 30 30	ACANE ACYET GOSH
41		RED EARTH	GN2.12	15 15 19 20 20	6.5 6.5 7. 7.5 7.5	KH	24 24 26 30 30	ACANE EUPOP SSHT
96		RED EARTH	GN2.12	15 15 19 20 20	6.5 6.5 6.5 7. 7.	H	30 30 30 30 30	EUPOP - GWLN
48		RED EARTH	GN2.12	15 15 19	6.5 5.5 6.	KH	30 30 30	ACANE ERGIL SOWL
97		RED EARTH	GN2.12	15 15 19	6.5 7. 7.5	H	30 30 30	ACANE - GOST
139		RED EARTH	GN2.12	15 15 19	6.5 6. 6.	KH	30 30 30	ACANE - FOSH
69		RED EARTH	GN2.12	15 15	6.5 6.	K	23 23	ACANE ERLON GOST
110		RED EARTH	GN2.12	15 15	5.5 5.5	H	30 30	ERNUC ACANE SOHT
124		RED EARTH	GN2.12	15 15	6.5 7.	KH	30 30	ACANE TRCOL FOSH
147		RED EARTH	GN2.12	15 19 19 19 19	6. 6. 6. 7.5 7.5	KH	30 30 30 30 30	ACANE ERGIL SOST
71		RED EARTH	GN2.12	15 19 19 19 20	6.5 6.5 7. 7. 7.5	K	30 30 30 30 30	ACANE EUPOP SSHT
109		RED EARTH	GN2.12	15 19 19 19	6.5 7. 8. 8.5	KH	30 30 30 30	ACANE ARCON GOSH
50		RED EARTH	GN2.12	15 19 19	6.5 6.5 8.	KH	30 30 30	ACANE - OST
79		RED EARTH	GN2.12	15 19 19	6.5 6.5 6.5	KH	30 30 30	ACANE CAART SOST
148		RED EARTH	GN2.12	15 19 20	5.5 6. 7.	K	30 30 30	ACANE ERPER GOSH
7		RED EARTH	GN2.12	19 14 14 15 15	6.5 6.5 6.5 6.5 6.5	KH	26 26 26 26 26	ACANE - GSHH
93		RED EARTH	GN2.12	19 19 19 19 21	5.5 6.5 6.5 6.5 6.5	H	26 30 30 30 30	ACANE - FOSC
60		RED EARTH	GN2.12	19 19 20 20 21	6.5 6.5 6.5 6.5 7.5	CH	24 30 30 30 31	ACANE EUPOP GWLL
42		RED EARTH	GN2.12	19 20 20	5.5 5.5 5.5	KH	30 30 30	ACANE - GSHT
194		RED EARTH	GN2.12	5 5 8 14 17	7. 7. 7. 6.5 7.5	LK	30 30 30 30 23	TRBAS WASPP SSHU
12		RED EARTH	GN2.12	6 6 6 7 9	6.5 7. 7. 7. 7.	L	30 30 30 30 30	ACANE ERERI GOST
88		RED EARTH	GN2.12	6 6 9 14 14	7. 7. 7. 7. 7.	L	30 30 30 30 -	ERSTU DOATT SSHN
90		RED EARTH	GN2.12	6 6 9 9	6.5 7. 7. 8.5	L	30 30 30 30	ACANE EUTER FOSH
106		RED EARTH	GN2.12	8 8 9 9 14	6. 6. 6. 6. 6.	LH	30 30 30 30 30	ERERI ERHEL SHLT
77		RED EARTH	GN2.12	8 9 9 14 14	5.5 5.5 5.5 7. 7.		30 30 30 30 30	ACANE ERERI GOST
91		RED EARTH	GN2.12	8 14	6. 6.	H	30 30	ACSPP - FOSH
108		RED EARTH	GN2.12	9 15 15		H	30 30 30	ACANE SAKAL OSH
29		RED EARTH	GN2.12	9 9 14 14 19	6.5 6.5 6.5 6.5 6.5	K	30 30 30 30 30	ACANE ERGIL SSHT
32		RED EARTH	GN2.12	9 9 14	6.5 6.5 6.5	KH	30 30 30	EUPOP ERSTU SOWL
93		RED EARTH	GN2.12	9 9 9 14 14	7. 7. 7. 7. 7.	LH	30 30 30 30 -	SAKAL WASPP WLH
99		RED EARTH	GN2.12	9 9 9 14 14	6. 6. 6. 6.5 7.	H	30 30 30 30 30	ACANE ERBOW SOSH
38		RED EARTH	GN2.12	9 9 9 9 14	6.5 6.5 6. 6. 6.	K	30 30 26 26 26	ERERI ACANE SHLT

criterion and tabulated. An example of this was the sorting of sites according to the degree that they were affected by wind and water erosion. Costs of extracting this information were in the vicinity of $4.50 manually (one man-hour) vs $3.00 by computer.

The listing of sorted data also allowed the preparation of tables for soil laboratory data for each of the major soil groups. The cost of preparing all the soil analysis tables in the WARLUS report would have been in the vicinity of $420 if done manually. As it was actually carried out the cost was approximately $50 for computer time plus $65 for supplementary manual processing. Since then this procedure has been fully automated with further savings in man-hours in later surveys. Presence and frequency of occurrence of vegetation species for all sites, specified groups of sites or individual sites were easily tabulated at considerable savings. As an example, an inversion (which indicates frequency of occurrence and site location for classes of a specified attribute) about the five most common vegetation species cost $62 manually as against $16 by computer. An inversion about species (complete species list) would have been 2 to 3 times this cost manually whereas it cost only $8 for computer time. This type of processing allowed positive statements to be made about vegetation. In some cases these useful displays would not have been obtained if the data had had to be handled manually.

It was also possible to request more sophisticated sorting and classification. Table 2 shows one example of a multiple sort, costing approximately $3. The attributes involved here had previously been sorted manually for presentation in a paper that was required before the program operating on the computer file was operational. Approximately one week was spent organizing the data using manual techniques; wages for this period would have been approximately $250. Up to 40 sorts of this nature could have been produced by the computer for the same cost but it is very unlikely that this number would have been necessary to arrive at the particular table finally used.

The use of simple sorting and formatting facilities was not only cheap but led to improved and more accurate descriptions of the soil, vegetation and land types. It also provided a check on any initial bias in the grouping, and in some cases selection, of sites. In addition to this improved job satisfaction, as the extraction and tabulation of data is an onerous task which most people avoid as far as possible.

A number of other types of data manipulation were not used in the preparation of the initial WARLUS report (Anon. 1974) but have been used subsequently by others interested in data in the WARLUS file or other titles outside the Department of Primary Industries. One such is the use of the INFOL formatting facility to prepare facsimilies of the traditional CSIRO Soils Division profile description sheets, i.e. fully decoded descriptions of sites. This was done on an experimental basis for the WARLUS file and the Soils Division Solodic Soils file (SOLSTUD) (W.T. Ward priv. comm.) at a cost of 20c per site. Such output could be used directly in reports without

further human intervention, thus eliminating further transcription errors.

A second type of data manipulation involves selecting appropriate data and passing it to a program which maps site locations and attribute values on the Australian Map Guid. For example, the sites carrying sandalwood (Eremophila mitchellii) in the WARLUS survey were incorporated into the Queensland sandalwood study (Beeston priv. comm.) and were mapped (scale 1:6,000,000) at a cost of $1.25 for 146 sites, using a FORTRAN program. Similarly, specified sets of data have been extracted for pattern analysis and other multivariate statistical analyses.

Conclusion

The value of using computer assistance can be seen from the costs presented above. Despite the fact that this was an experimental study and involved development costs not involved in subsequent surveys, the computer-oriented system proved to be an efficient method of collecting, storing and retrieving data. Two major problems had to be overcome to achieve this. In the first place, the poor operational record of INFOL restricted its use at critical stages of interpretation and in the long run forced the development of alternative programs The second problem was the initial resistance to the use of computers by field and laboratory staff. Once the problems were overcome efficient use followed.

Initial feelings that the computer-oriented system and particular sets of data might have a short life (up to four or five years only) have proved to be both right and wrong. Whilst INFOL has now been replaced by other programs the data collected in 1970-71 is still of value and being constantly used both by the original collectors and, more importantly, by others. There is no doubt that there is much freer dissemination of data from the WARLUS survey because of its ready accessibility via the computer, than from earlier surveys.

The use of a computer-oriented information system in WARLUS also improved job satisfaction by avoiding onerous jobs, increased accuracy, increased the amount of interpretation of collected data and raised the standard of the final publication.

References

Anon. (1974). Western Arid Region Land Use Study - Part 1. Tech. Bul. No. 12, Division of Land Utilization, Department of Primary Industries, Brisbane. pp. 132 + cxviii.

Dawson, N.M. (1972). Land Inventory and Technical Guide, Miles Area, Queensland - Part 1: Land Classification and Land Use. Tech. Bul. No. 5, Division of Land Utilization, Department of Primary Industries, Brisbane. Pp. 64 + xxxx.

Dawson, N.M. (1976). Collection, storage and retrieval of rangeland resource data. (In press.)

Moore, A.W., W.T. Ward and C.H. Thompson (1974). Computer storage of soil data: use of a generalized file management system. Trans. tenth int. Congr. Soil Sci. 6:684-691.

Norris, J.M., S.W. Cumpston and J.R. Sleeman (1971). A storage and retrieval system for micromorphological descriptions of soils. CSIRO Div. Soils Tech. Paper No. 10.

Sibley, E.H. (1974). Data management systems - user requirements. In "Date Base Management Systems" (ed. D.A. Jardine), pp. 83-104 North-Holland Pub. Co., Amsterdam.

Walker, J., D.R. Ross and G.R. Beeston (1973). The collection and retrieval of plant ecological data. CSIRO, Woodland Ecol. Unit Pub. No. 1.

Participants to the meeting of the working group on soil information systems, Canberra, A.C.T., 2 - 4 March 1976

(From Australia if not otherwise indicated)

Mr D.W. Armstrong
Soil Survey Section, Department of Agriculture, P.O. Box 618,
NARACOORTE S.A. 5271

Dr J.W. Bowden
Department of Agriculture,
2 Jarrah Road,
SOUTH PERTH W.A. 6151

Dr G.P. Briner
Department of Agriculture,
Box 4041, G.P.O.,
MWLBOURNE VIC. 3001

Dr P.A. Burrough
School of Geography, University of New South Wales,
P.O. Box 1,
KENSINGTON N.S.W. 2033

Dr J.D. Colwell
CSIRO Division of Soils,
P.O. Box 639,
CANBERRA CITY A.C.T. 2601

Mr B.G. Cook
CSIRO Division of Land Use Research, P.O. Box 1666,
CANBERRA CITY A.C.T. 2601

Mr R.S. Cormack
Department of Primary Industries, Meiers Road,
INDOOROOPILLY QLD 4068

Mr N.M. Dawson
Department of Primary Industries, Meiers Road,
INDOOROOPILLY QLD 4068

Dr E.G. Hallsworth
Chairman, CSIRO Land Resources Laboratories, c/- CSIRO Division of Soils,
Private Mail Bag No. 2,
GLEN OSMOND S.A. 5064

Dr R. Lee
Soil Bureau, DSIR, Private Mail Bag,
LOWER HUTT, NEW ZEALAND

Mr L.G. Lynch
Soil Conservation Service of N.S.W., Box R201 Royal Exchange P.O.,
SYDNEY N.S.W. 2000

Mr I.H. Lynn
National Water and Soil Conservation Organization,
Ministry of Works and Development, P.O. Box 1479,
CHRISTCHURCH, NEW ZEALAND

Dr A.W. Moore
CSIRO Division of Soils,
Cunningham Laboratory, Mill Road,
ST. LUCIA QLD 4067

Mr W. Papst
Soil Conservation Authority of Victoria, 378 Cotham Road,
KEW VIC. 3101

Mr G.H. Price
Consolidated Fertilizers,
P.O. Box 140,
MORNINGSIDE QLD 4170

Mr A.M.H. Riddler
N.S.W. Department of Agriculture, Private Mail Bag 10,
RYDALMERE N.S.W. 2116

Mr R.K. Rowe
Soil Conservation Authority of Victoria, 378 Cotham Road,
KEW VIC. 3101

Mr J.R. Sleeman
CSIRO Division of Soils,
P.O. Box 639,
CANBERRA CITY A.C.T. 2601

Dr J.L. Smith
CSIRO Division of Computing Research, P.O. Box 1800,
CANBERRA CITY A.C.T. 2601

Mr K. Stackhouse
Department of Agriculture,
P.O. Box 46,
LAUNCESTON STH., TAS. 7250

Mr K.A. Styles
Soil Conservation Service of N.S.W., Box R201 Royal Exchange P.O.,
SYDNEY, N.S.W. 2000

Mr G. Tregenza
Department of Environment, Housing & Community Development, P.O. Box 1890,
CANBERRA CITY, A.C.T. 2601

Mr A. Webb
Department of Primary Industries, Biloela Research Station, P.O. Box,
BILOELA, QLD 4715

Mr K.G. Wetherby
Soil Survey Section,
Department of Agriculture,
P.O. Box 156,
CLEVE, S.A. 5640

Mr I. Wong
Department of Agriculture,
Kuala Lumpur
MALAYSIA