74-63

canadian centre for geoscience data
centre canadien des données géoscientifiques

GEOLOGICAL SURVEY
PAPER 74-63

COMPUTER-BASED SYSTEMS FOR GEOLOGICAL FIELD DATA

An international state-of-the-art review for 1973 conducted by COGEODATA, International Union of Geological Sciences, in collaboration with the Division of Earth and Environmental Sciences, UNESCO and the Canadian Centre for Geoscience Data.

W. W. HUTCHISON
Chairman, COGEODATA

Published by the Geological Survey of Canada
Department of Energy, Mines and Resources, Ottawa

1975

Energy, Mines and Resources Canada
Énergie, Mines et Ressources Canada

GEOLOGICAL SURVEY
PAPER 74-63

COMPUTER-BASED SYSTEMS FOR GEOLOGICAL FIELD DATA

An international state-of-the-art review for 1973 conducted by COGEODATA, International Union of Geological Sciences, in collaboration with the Division of Earth and Environmental Sciences, UNESCO and the Canadian Centre for Geoscience Data.

W. W. HUTCHISON
Chairman, COGEODATA

1975

Available by mail from Information Canada, Ottawa, K1A 0S9
and at the following Information Canada bookshops:

HALIFAX
1683 Barrington Street

MONTREAL
640 St. Catherine Street West

OTTAWA
171 Slater Street

TORONTO
221 Yonge Street

WINNIPEG
393 Portage Avenue

VANCOUVER
800 Granville Street

or through your bookseller

Price: Canada: $4.00
Other Countries: $4.80 Catalogue No. M 44-74-63

Price subject to change without notice

Information Canada
Ottawa, 1975

CONTENTS

FOREWORD

The Committee on Storage, Automatic Processing and Retrieval of Geological Data, known more commonly as "COGEODATA", is a committee of the International Union of Geological Sciences. One of the purposes of COGEODATA is to appraise existing systems for storing geological and allied data. In an effort to meet this need COGEODATA convened a 5-day seminar at UNESCO headquarters, Paris, France, 5-9 November 1973, to discuss the use of computer-based techniques in geological field work and geological data-base management. With the assistance and cooperation of the International Geographical Union and UNESCO's Division of Earth and Environmental Sciences, 41 invited participants attended, representing nearly all countries with significant achievement in this field. Included also were observers representing other organizations having an interest in the topic, including CODATA (International Council of Scientific Unions), the International Association for Mathematical Geology, and UNISIST (World Science Information System). This publication summarizes results of the seminar.

Each of the 18 contributions is a short summary of developments, concepts and plans reported at the seminar. For additional details readers are refered to the bibliographies following each paper and they are encouraged to correspond with the authors, whose complete mailing addresses are provided. Illustrations have been reproduced as received, in order to help expedite publication. COGEODATA acknowledges with thanks the additional contribution received from J.M. Botbol and R.W. Bowen, United States Geological Survey, who were not able to present their paper at the seminar.

In keeping with the spirit of this report, readers may be interested to know that authors' typescripts were entered into a computer-based text-editing system for processing at the Canadian Centre for Geoscience Data, and subsequently set in type by a computer-driven photocomposition system (Alphatext Systems Limited, Ottawa). This is the second paper published by the Geological Survey of Canada to be so processed (the first being COGEODATA's bibliography for 1970-72, GSC Paper 73-14) and I trust this experience augers well for the future application of modern information technology to geoscience.

Canadian Centre for Geoscience Data
Ottawa, Canada

C.F. Burk, Jr.
Secretary, COGEODATA

Introduction to geological field data systems and generalized geological data management systems

W. W. Hutchison, Chairman of COGEODATA
Geological Survey of Canada
601 Booth Street
Ottawa, Ontario K1A 0E8
CANADA

During much of its existence, COGEODATA has devoted considerable energy to the standardization of geological data and the formalization of procedures for handling geological data. The series of papers which follow is, in a way, a departure from this trend but aims at one of the objectives of COGEODATA, namely: "an appraisal of existing systems for storing geological and allied data". The reason for following this theme is as follows: regardless of how well we standardize, we must also establish how facile we will be in handling the data and at the same time discover empirically which appear to be 'natural' standards. Geological field data systems were chosen for the following reasons: the amount of activity in the area during the past 5 years, the lack of any effective communication among those researchers, and the fact that field data are fundamental to all geological information.

The meeting in Paris was originally planned following my appointment as Chairman of the Field Data Working Group of COGEODATA in 1972. The intention was for a small meeting of a working group to discuss the activities in this particular field. After some discussion, it was decided to bring together as many workers as possible in this field, because of the diversity of geological field mapping, and furthermore the diversity of systems that have been developed to cope with collection of field information. Moreover, if geological data are to have any meaning, then they must be related to the field setting and, accordingly, systems developed for the collection of geological field data fundamentally affect how we use the data derived during this initial exercise. Also, UNESCO kindly arranged for experts in spatial data systems of the International Geographical Union to participate in this meeting; their contribution and advice were most valuable.

The objectives of the meeting, held at Unesco headquarters, Paris, 5–9 November 1973 were:

1. To determine the effectiveness of computer-based systems developed for geological field data,
2. To determine which types of geological terrain are most amenable to mapping, using computer-based techniques,

3. To analyze scientific procedures and their adaptation to computer-based systems,
4. To establish the nature of constraints versus the degree of freedom in current computer technology,
5. To inform users and potential users of the state of the art, and
6. To provide a forum for communication among a group of workers.

It was decided to add the topic of generalized geological data management systems to this program because of the major hiatus that has existed between computer scientists and geologists. In an effort to reduce this hiatus, it was decided that some experts in data management systems in geology should participate so that they could perhaps better appreciate some of the needs of geologists. Conversely, the field geologists should perhaps understand better the concepts and techniques of current data management systems. It should be stressed at this point that there is no implication that a generalized data management system cannot be used in geological field mapping. Nevertheless, very few field systems that have been developed utilize generalized systems.

Aim of this Publication

The aim of this publication is not to provide a detailed account of the meeting. Instead, it is a collection of 'vignettes' highlighting some aspects discussed at the Paris meeting. The reason for this is simple. A booklet with short, to-the-point comments is more likely to be read. Furthermore, most of the systems, models, and concepts discussed in Paris have already been published and references to these publications are contained in the bibliographies. Accordingly, the main aim of this publication is to communicate swiftly to users and to potential users, techniques and concepts currently in use, and in this way to act as a directory so that the user or potential user may contact that particular research worker directly. The state-of-the-art is evolving too quickly for us to pause and drain or divert substantial energy to a major comprehensive publication.

Historical Development

The First Wave. When COGEODATA first came into being, geological data were considered in terms of digits (in numerical or alphanumerical form). For geological data to be stored, retrieved and communicated, they had to be coded. Accordingly, initial efforts naturally stressed standardization and normalization of procedures to handle geological data. To begin with, this showed great promise so that groups in industry (especially those in oil exploration and in mining production) very quickly adapted certain techniques for practical use. Large data management systems were quickly developed in some companies and to a lesser extent, in some government agencies, but a number of these grew and then, like the end of the dinosaurs, stagnated or collapsed. An analysis of why this happened could be exceedingly beneficial to the users and potential users currently appearing on the market. Some factors include: Lack of clear objectives in establishing a computer-based system; artificial constraints imposed by the system; and high costs and poor planning of documentation, editing and updating. Those systems that survived this first wave commonly indicate they were designed for specific purposes with a very clear idea of the problems and objectives.

The New Wave. During the last few years a number of practical operating systems for recording geological field data have appeared. Coupled with this, there is the fact that computer technology no longer requires us to be so constrained. We are no longer required to code our data and information so that it may be 'squeezed' through a narrow pipe. Instead the 'diameter' of the medium has vastly increased to the point in fact that any 'normal' field notes can now be economically typed into the computer (either directly or via tape or punch-card). These two points are very important because they stress the fact that practicality has arrived in data capture and furthermore, as time goes on, the degree of freedom with which to analyze our data or information is quickly increasing.

Activity during the first wave was exceedingly beneficial. It fostered the growth of groups which set out initially to 'mimic' (using computer-based systems) the techniques which have been used by scientists in geological mapping over the last 50 years. At this point, I arrive at what I think is perhaps the most crucial juncture in this whole matter of geological data, computers and the science of geology itself, namely: the hierarchy of the data. In the past, comment has been made about the failure of the field geologist to record data systematically. Analysis of this comment brings one to realize that geological field data are often regarded as point sets of data as though they have been recorded digitally by some blind machine. The geologist does not stop at one point A, look down at his feet and record all significant data within a radius of three metres of his feet, then close his eyes and walk blindly onwards for a hundred metres or one kilometre and then stop, look down, open his eyes at point B and record all data within a radius of three metres. Instead he is more concerned in the first instance in establishing the relationship between data set A and data set B, and recognizing (using his own inboard computer) whether or not there is a difference, and if so, what its nature might be. Each hour and each day in the field is spent working essentially interactively with the rock patterns to build up a picture of the field setting. A highly skilled field geologist is perhaps one of the most discriminating and sensitive pure scientists. I would contend, therefore (Hutchison, 1973) that there has been a general failure to recognize that geological contacts are prime data located on geological field maps (they can easily be represented as a "stream" of x-y-z coordinates in the computer), whose nature results from observation of the relationship between data set A and data set B. On this basis, 'spot' data within units are therefore of a lower rank than the data for contacts between units. Systems which have been most successful are those used in terrain where contacts between units are not exposed or appear to be gradational.

The importance of capturing geological contacts is that they can then be utilized for production of maps (this has been done for a number of years at the Royal College of Art, Bickmore and Kelk, 1972 and Thorpe, this paper), for cartographic analysis and as a basis for a data management system.

Outcome of the Paris Meeting

Emerging from the Paris meeting was the appearance of three groups: one using the 'medium', where appropriate, as a means toward an end, the other in examining the 'medium' itself for handling geological data, and the third interested in both areas. As a consequence, all participants were not focused on the same research targets. Nevertheless, this divergence had a tempering influence, whereby geologists with diverse objectives could be informed of techniques currently being appraised, while researchers in data management systems could better appreciate geologists' needs.

Major contributions to advances in use of computer technology span the spectrum from the requirments of the scientists (see contribution by de Heer and Bie and also Gordon), to the production of final colour maps such as discussed by Thorpe. Within this spectrum we have contributions on some of the more successful field data systems, for example: GEOMAP – Berner, Ekström, Lilljequist, Stephansson, and Wikström; Canadian Systems – McRitchie; Finnish systems – Pipping; DASCH – Vinken; and SARS – Hutchison and Roddick. Generalized systems for laboratory and field data include Aarhus University System II – Platou, and the CRPG System – Grandclaude. The chief attributes of all these systems is the ability to systematically (implying some 'standards') record, store and retrieve large volumes of geological information so that selection may be analyzed by application programmes including plotting. Problems identified are the limited range of applicability of field systems (most are designed for specific terrain – usually metamorphic or plutonic terrain in which well defined stratigraphy is not readily identifiable), the hiatus between geologist and computer, the need to formulate a workable and consistent methodology, and, where a high degree of success is encountered (such as GEOMAP), acquiring money and manpower to establish that system for a 'beginner'.

Generalized data management systems include SAFRAS by Sutterlin, SIGMI by Kremer, Lenci and Lesage and G-EXEC by Jeffery and Gill. There are many more data management systems designed for or adapted to use in geology. However, the three above are interesting for a variety of reasons. For example, SAFRAS was the first portable DMS using free format allowing for repeating record types. It has a particularly powerful retrieval capability but the COBOL programs may hamper portability because of machine dependence.

SIGMI is a more sophisticated extension of the SAFRAS concept, allowing for a higher degree of 'nesting' in the hierarchy. In contrast with SAFRAS and SIGMI, G-EXEC is a generalized FORTRAN system which has been built as a series of modules. It is probably as portable as any generalized system may hope to be with respect to fundamental differences between computers. In terms of a small, comprehensive system, it has many attractive features, not the least of which is the varied yet heavy workload it has carried to date.

Again I would like to stress that there are many other geological data management systems as well as some spatial data systems which could easily be adapted for geological use. This publication is not intended to provide a list of such various systems. A potential user in selecting a system should, however, consider planned use of data, the constraints of the make (or makes) of computer a scientist may use, the operating system on that computer, financial budget, systems and programming staff available, and number of scientists to be supported. Also to be considered is the question of whether scientists will use computers largely as 'accounting' machines for routine computation and plots (batch mode), or whether they would wish to have the freedom to work on an instantly accessible basis to do 'research' on their data (interactive mode). I am convinced that on-line, interactive capabilities are (and will become) essential for certain research endeavours – yet again such a requirement puts a further constraint on the data management system which might be used.

Conclusions

In a general discussion there was a diversity of opinion concerning greatest needs for future effort. These reflect the diverse problems currently being encountered by the broad spectrum of participants.

There was general agreement to a proposal to have COGEODATA seek funds to sponsor an evaluation of operational field data systems. This matter will be followed up.

At the 'systems' level, there was interest in a codicil for data structure, for establishing a range of 'system' components, for increased collaboration on system development, evaluation of systems for data management and establishing standards for transfer of data.

At the 'data collection' level, there is a clear need to further reduce the gap between field geologist and computer, for greater systematization in both 'standards' and procedures in collection, and to reduce the constraints of current systems.

At the level of those not familiar with computer techniques in geology, there is a clear need for some education, and for participation of COGEODATA in activities of CODATA (ICSU) and IGCP (International Geological Correlation Programme).

Research on all these topics will be done, some perhaps under the aegis of COGEODATA. Each area, in its own right, demands research, but limited budgets and manpower dictate that priorities be set.

The Future

Readers will probably have been waiting for some sort of analysis of field data systems in terms of components, costs, applicability and effectiveness. Some data on these items are contained in the appendices. In spite of the tremendous progress made by the pioneers in field data systems, we still have a long way to go towards making such systems more widely applicable. At present, I would have little hesitation about strongly recommending use of a computer-based system in a long term mapping project in a metamorphic/granitic terrain where stratigraphy is largely lacking or obscured (by over-burden), and where many geologists will be employed.

For smaller, shorter term projects, and where stratigraphy is well defined, current techniques may be of limited assistance, requiring too much effort in implementation relative to compiling energy available. I regard a computer-based system as simply a medium, and it is for the field geologist to decide whether such a medium suits his work style better or has an advantage over pencil and paper. This assumes that the geologist's notes are simply a temporary project file. If, however, it is decided that notes will be archived, this could be an added reason for a computer system. At present, I would be concerned that there could be energy loss through handicapping the geologist in the field, followed by additional effort required for effective curation of the field notes.

Field data files may quickly grow and with large files we tend to slip into one of the major pitfalls, namely: the amount of energy required for file management, editing and updating a large file tends to consume so much time that more time is spent managing the data than attaining the ultimate objective. Another pitfall is that a standard may have been prematurely established and imposes a stultifying effect.

Especially in field geology we must approach the matter of standards with care. Obviously each field data system is in a sense a closed system, and one would hope that standards used internally are consistent. Most standards (aside from those dealing with location) have nothing to do with computers. They are instead geological problems requiring resolution by geological experts in their particular field using data hierarchy and classification. A benefit from field data systems is that we can test proposed standards and present others which have naturally appeared or 'evolved'.

At this point, I return to one of the comments I made earlier, that an excellent field geologist is "one of the most discriminating and sensitive pure scientists" in this age of technology. The challenge, as I see it, is for the computer scientist to quietly follow such a field geologist and determine how computer techniques could facilitate the information collection. I feel that this is an area where much will be achieved in the near future. Such examples are new techniques for capturing data on tape at costs lower than normal typing through use of off-line terminals.

Digitization not only of point data but also of contacts, allows us to put the main elements of a map into the computer and thus potentially available for cartographic analysis. Of particular importance here, is that statistical and other analyses are not 'one way' synthesis but can be fed back, as part of an iteration, to allow evaluation of domain boundaries. In this way data may be given geological significance and as domain boundaries are changed, then so must the geological context of all contained data. Many geologists do this every day and it is essential that any computer system for field geology must have the same capability.

An example of what may become commonplace in the future is the contribution of Smedes, USGS, which demonstrates how a modelling-mapping system combining remote sensing, natural resource, social, economic and cadastral data has been applied to planning and decision-making.

I have no doubt in my own mind that computer techniques will give us a degree of freedom that we never had before. Through this we will be able to make a contribution to the science of geology and its application. However, I do not see this as an easy process that will develop quickly, and neither will it be comprehensive. Instead I anticipate that setbacks, frustration and higher than anticipated costs will be commonly encountered in development. For this reason, I would like to pay particular tribute to the participants of the Paris meeting. Through their efforts and determination, substantial hard-earned progress has been made, not only in accomplishment, but in discoveries of pitfalls. These pioneers have a wealth of experience. Thanks to the support of IUGS (through COGEODATA), and of UNESCO, the participants were brought together in one place to minimize duplication of effort required for future progress. This meeting was essential and valuable and I would suggest that the reader seeking further information should contact directly any contributor, or write to the COGEODATA Secretary, Dr. C. F. Burk, Jr.

Acknowledgments

I would like to acknowledge the vital assistance of Dr. C. F. Burk, Jr., our Secretary, and National Coordinator of the Canadian Centre for Geoscience Data, in planning the Paris meeting, and in editing and typesetting (appropriately, with the aid of a computer-based text editing and photocomposition system) the contents of this publication. Furthermore, I thank the Geological Survey of Canada for publishing the contributions in this paper.

Particular thanks must be given to Dr. K. Lange and his Secretary, Mme. Espinasse, who made arrangements to hold our meeting at UNESCO.

At the meeting, Drs. O. Stephansson, M. Kremer, W. D. McRitchie, and G. D. Williams bore a substantial responsibility as Chairmen in fostering and guiding the many valuable discussions.

I wish to thank UNESCO and the International Geographical Union (IGU) for the participation of: A. R. Boyle, Saskatoon, Canada; H. W D. F. Marble, Buffalo, U.S.A.; W. R. Tobler, Ann Arbor, U.S.A.; and R. F. Tomlinson, Ottawa, Canada.

Above all, I thank the participants, because it is these pioneers whose efforts, sacrifices and determination have contributed so much to the progress that has been made to date.

Reference

Hutchison, W. W.
1973 GSC computerizing system for map production, analysis: Northern Miner, v. 59, no. 37, p. 95.

PART 1: GEOLOGICAL FIELD DATA SYSTEMS

Introduction

W. D. McRitchie
Department of Mines, Resources and
Environmental Management
Geological Survey Section
993 Century Street
Winnipeg, Manitoba R3H 0W4
CANADA

For several years now many geologists have recognized the need for guidelines formulating a workable and consistent methodology for establishing data files that can be processed with a high degree of reliability. Many might claim that such guidelines are inherent in the efficient applications of existing geologically oriented field data gathering procedures. In practice, however, it is often found that standards vary or are not defined, inconsistencies exist both within and between files, and more often or not the data required have not been collected or recorded.

The following papers outline some of the initial attempts that have been made by geologists from many different countries to upgrade the consistency and reliability of geological data-gathering procedures in the hope that the data files so generated might be processed with an equal degree of meaningfulness and reliability.

One of the greatest limitations that most geologists are forced to work under is a lack of time. Time to collect the information they need, and time to work with the raw data, correcting, sifting, and analyzing it, searching for meaningful covariances which will lead them to resolve the problems they are confronted with. If the content of their data files is consistent, then the greatly enhanced capability for testing multiple working hypotheses and gaining a feel for the data through machine processing can lead to a more rigorous and profitable conceptual analysis of the perceptually derived data.

GEOMAP

H. Berner, T. Ekström, R. Lilljequist, O. Stephansson and A. Wikström
Uppsala Universitet
Geologiska Institutionen
Box 555
S-751 22 Uppsala 1
SWEDEN

Introduction

Increasing recognition of the ability of data processing systems to handle large amounts of current and potential geological data suggests the extension of data processing methods into geological mapping. For this reason the field data storage and retrieval system GEOMAP was developed in 1969-70. The system has been used for four field seasons in Scandinavia and was slightly modified in 1972. A meaningful use of computers calls for systematic recording of consistent observations. To fulfill this condition GEOMAP utilizes a standard field data sheet which serves as a checklist. Information not covered by the data sheet may be recorded as processable notes. The combination of a standard sheet with processable notes gives GEOMAP a high degree of flexibility.

The data base is linked to a data management system which contains a set of retrieval programs and routines for standard statistical methods.

Field Data Sheet

Observations and measurements made on an outcrop are subdivided with reference to different rock types. Information about a rock type can be given on more than one sheet (Figure 1). The description includes:

1. Coding on the input document
2. Referenced notes on the back of the input document
3. Common notes on the back of the input document.

Only a few remarks about the field data sheet are made here; a special guide for the user is available upon request.

The sheet is divided into segments, for example, location and references, petrology, tectonics. Any orthogonal coordinate grid may be used for indicating the geographical location. If more detailed

GEOMAP

MAP UNIT: DATE:

A1953
A1954

N-CO-OR	E-CO-OR	ALTITUDE	GEOLOGIST	OUTCROP NO.	SERIAL NO.
6 7 8 9	10 11 12 13	14 15 16 17	18	19 20 21 22	23

TYPE OF OUTCROP

124

OUTCROP $< 5\ m^2$ . . . . . A
OUTCROP 5-100 m^2 . . . . B
OUTCROP 100-2000 m^2 . . C
CONTINUOUS OUTCROP . . D
ROAD CUT <10 m . . . . E
ROAD CUT 10-50 m . . F
QUARRY . . . . G
MINE, PROSPECT . . . . H
BORE-HOLE . . . . K
KEY-OUTCROP L
UNCERTAIN OUTCROP. M
BOULDER . . . . N
PILE OF BOULDERS . P

REFERENCE

25

TECTONICS AS PRECED. OBSERVATION . . . 1
PETR. AND TECT. AS PRECED. OUTCROP . . 2
PETROGRAPHY AS PRECED. OUTCROP . . 3
TECTONIC AS PRECED. OUTCROP 4
PETR. AND TECT. AS SECOND LAST OUTCR. 5
PETROGRAPHY AS SECOND LAST OUTCR. 6

OBJECTS

TYPE I: 126

PHOTO 1
SKETCH 2
PHOTO + SKETCH . . 3
DESCRIPTION 4
PHOTO + DESCRIPT. . 5
AREA DESCRIPT. . . 6
OTHER 7

TYPE II: 127 — OBJECT NO: 28 29 30 31 32 33

SPECIMEN 1
SPECIMENS 2
CHEMICAL ANALYSIS 3
SPEC.+CHEM. ANALYSIS 4
MICROSCOPIC ANALYSIS 5
SPEC.+MICR. ANALYSIS 6

PETROGRAPHY

I: 34 135 II: 36 137

STRUCTURE

SEE CHART

LAYER THICKNESS

138

<2 mm . . . 1
2-20 mm . . . 2
2-20 cm . . . 3
2-20 dm . . . 4
2-20 m . . . 5
> 20 m . . . 6
VARIABLE . 7

COLOUR

39 140

SEE CHART

GRAIN SIZE

mm — 141

<0.05 . . . 1
0.05-0.3 2
0.3-1.0 3
1.0-1.5 4
1.5-3.0 . . 5
3-5 6
5-30 7
> 30 8

TEXTURE

142

EQUIGRANULAR 1
INEQUIGRANULAR 2
PORPHYRITIC 3
OPHITIC 4
PORPHYROBLASTIC . . . 5
GRAPHIC 6
CLASTIC 7
OTHER 8

MEGAGRAIN

MINERAL: 43 144

SEE CHART

HABIT — 145

ANGULAR, ELONGATED 1
ANGULAR, EQUIDIMEN 2
ANGULAR, ROUNDED 3
LENTICULAR 4
ROUNDED 5
VARIABLE 6
OTHER 7

SIZE mm — 146

<2 1
2-5 2
5-10 3
10-20 4
20-40 5
>40 6

QUANTITY % — 147

<1% . . X
1-5% . . Y
10% . . 1
20% . . 2
30% . . 3
40% . . 4
etc.

MOBILISATE

TYPE — 148

PEGM. WHITE-GREY . . . 1
PEGM. WHITE-RED 2
INEQUIGR. WHITE-GREY. 3
INEQUIGR. WHITE-RED . 4
GRANITIC WHITE-GREY. 5
GRANITIC WHITE-RED . 6
GRANODIORITIC 7
DIORITIC 8
OTHER 9

QUANTITY % — 161 (AS 147)

MINERALOGY

	I TYPE	I QUANT.	II TYPE	II QUANT.	III TYPE	III QUANT.	IV TYPE	IV QUANT.
	49 150	151	52 153	154	55 156	157	58 159	160

ROCK

STRATIGRAPHY	TYPE	MINERAL, ALTERATION	QUANTITY %
62 63 154	65 66 167	68 169	170 (AS 147)

OTHER OBSERVATIONS

71 172

TECTONICS ETC.

PLANAR STRUCTURES

SS-SURFACE 1
S-SURFACE 2
PLANAR STRUCTURES . 3
AXIAL PLANE 4
JOINT 5
SET OF JOINTS 6
FAULT 7
SHEAR ZONE 8
OTHER 9

TYPE	STRIKE	DIP	PHASE
224	25 26 27	28 29	230
233	34 35 36	37 38	239
241	42 43 44	45 46	247

JOINT FREQUENCY

NUMBER OF JOINTS/5m ⊥ STRIKE

231, 240, 248

2-3 1
4-6 2
7-10 3
11-20 4
21-30 5
>30 6

STRATIGRAPHICAL SEQUENCE

232

NORMAL 1
INVERTED 2
V { 0°-179°, TOP . . . 3
V { 180°-360°, TOP . . . 4

DIMENSION

0.01-0.05 . . . 1
0.05-0.1 2
0.1-0.25 . . . 3
0.25-0.50 . . 4
0.5-1.0 . . . 5
1.0-2.5 . . . 6
2.5-5.0 . . . 7
5.0-10 8
>10 9

SYMMETRY

SYMMETRIC 1
ASYM { Z <2/1 . . . 2
Z 2/1-4/1 . . . 3
Z 4/1-10/1 . 4
Z >10/1 . 5
S <2/1 . . 6
S 2/1-4/1 . . 7
S 4/1-10/1 . 8
S >10/1 . 9

LINEAR STRUCTURES

FOLD AXIS, OBSERVED . . . 1
FOLD AXIS, CONSTRUCTED. 2
MINERAL LINEATION . . . 3
S-INTERSECTIONS 4
LINEATION UNSPEC. . . . 5
LINEATION+FOLD AXIS . . . 6
LINEAR STRUCTURES . . . 7
SLICKENSIDES 8
OTHER 9

TYPE	AZIMUTH	PLUNGE	PHASE
249	50 51 52	53 54	255
260	61 62 63	64 65	266

FOLDS

CLASS 1A . . . . A
CLASS 1B (PARALLEL) . . . B
CLASS 1C C
CLASS 2 (SIMILAR) . . . 2
CLASS 3 3
CHEVRON 4
PTYGMATIC 5
INTRAFOLIAL 6
FLOW 7
OTHER 8

TYPE	WAVELENGTH	AMPLITUDE	SYMMETRY
256	257	258	259
267	268	269	270

DYKES ETC.

DYKE 1
SILL 2
FILLED JOINT 3
IGNEOUS CONTACT . . . 4
SEGREGATION 5
GLACIAL STRIATION . . . 6
OTHER 7

TYPE	STRIKE	DIP	PHASE
271	72 73 74	75 76	277

FILLING — 278

GRANITE GNEISS . . . A
MASSIVE GRANITE . . B
PEGMATITE C
APLITE . . . . D
AMPHIBOLITE . . . . . . E
DOLERITE . . . . F
QUARTZ . . . . G
QUARTZ+ORE . . . . . H
ORE . . . . . K
CARBONATE L
SKARN . . . M
OTHER . . . N

THICKNESS

<2 cm 1
2-10 cm 2
1-2 dm 3
2-10 dm 4
1-2 m 5
> 2 m 6

Figure 1. Field data sheet.

```
MAP UNIT KATRINEHOLM SW                             YEAR     1973          OUTCROP NO 6
N-CO-OR 704410  E-CO-OR 130430                      GEOLOGIST O STEPHANSSON
TYPE                ROAD CUT 10-50 M
----------------------------------------------------------------------------------
ROCK       A        QUANTITY  80 %
STRUCTURE 1         UNIFORM BEDDING
          2         SCHISTOSITY
GRAIN SIZE          0.05-0.3 MM                     TEXTURE  EQUIGRANULAR
LAYER THICKNESS     2-20 MM                         COLOUR   MEDIUM DARK GRAY (74)
ROCK TYPE           CARBONACEOUS SHALE                       STRATIGRAPHY SIL
MOBILISATE          <1% PEGM WHITE GRAY
S-SURFACE               N 15 E, 70 W                PHASE 1
SS-SURFACE              N 15 E, 70 W
FOLD AXES               S 20 W, 10                  PHASE 1
FOLD TYPE 1         CLASS 1 C                       SYMMETRY SYMMETRIC
WAVE LENGTH         0.05-0.1 M                      AMPLITUDE 0.05-0.1 M
FILLED JOINT            CARBONATE                   N 5 E, 90 E THICKNESS < 2 CM
PHOTO+SKETCH
NOTE(S):
OBJECT              126,3  SEE COMMON NOTE
SYMMETRY            259,1  SEE SKETCH+PHOTO
ROCK TYPE           167,   FOLDING IN CARBONATE RICH PARTS
COMMON NOTE         PHOTO OF FOLDS IN CARBONATE RICH PARTS, FILM 1-3
----------------------------------------------------------------------------------
ROCK       B        QUANTITY 20 %
STRUCTURE 1         SCHISTOSITY
          2         UNIFORM BEDDING
GRAIN SIZE          <0.05 MM                        TEXTURE  EQUIGRANULAR
LAYER THICKNESS                                     COLOUR   MEDIUM DARK GRAY (74)
MINERAL    1        80% CARBONATE
ROCK TYPE           MARBLE (58A)                             STRATIGRAPHY BER
S-SURFACE               N 70 W, 20 E                PHASE 3
PHOTO+SKETCH
NOTE(S):
OBJECT              126,5  PHOTO OF OUTCROP WITH OFFERDAL  CHURCH IN BACKGROUND
TYPE PLANARSTRUCT 224,2  STRONGLY CRUSHED IN WESTERN PART OF OUTCROP
```

Figure 2. List of decoded information from an outcrop. Notes are printed at the bottom of each description.

mapping is desired, serial numbers can be used for switching the registration of coordinates into a tenth or a hundredth of the ordinary mapping scale.

The petrographical and tectonic data are coded numerically or alphanumerically. Some codes are preprinted on the sheet, but codes for structures, colours, minerals, rock types and folds are given on separate charts which the geologist carries together with the sheets in the field. The contents of the charts as well as the libraries of codes and decoded information within the data program can easily be updated and replaced. This allows the use of different schemes of rocks, structures, minerals, etc., which also makes GEOMAP more adaptable to different types of geological milieu.

The colours of the rocks are defined by means of a standard rock colour chart designed for field use.

In the 'Mineralogy' segment four major or minor minerals may be recorded on each field sheet. The relative abundances of the minerals are given in the 'Quantity boxes'. For the description of 'Rock type' the entry 'Stratigraphy' is used for registering any litho-, chrono-, or biostratigraphic division or local name of the rock. The codes chosen by the geologist are processed by the computer just like any other parameter given on the sheet.

Up to now geologists in Scandinavia have used two different rock schemes. One is a genetic classification of Precambrian rocks used by the Mapping Department of the Swedish Geological Survey; the other is used by the Norwegian Survey and contains a classification of igneous rocks by Streckeisen.

In the tectonic segment of the data sheet linear and planar elements as well as dykes can be recorded for each rock type. Folds are coded according to the geometric classification of Ramsay.

The back of the sheet is used for notes and sketches. There are two types of notes in the GEOMAP system. Referenced notes are made to complete special information on the data sheet. These notes can be retrieved selectively with respect to any referenced observation from the data base. The general notes are only printed out together with all information about the rock types in the comprehensive outcrop list. These notes are used for more general descriptions and may have a considerable length.

Updating, Analysis and Retrieval of Data

The geological information is transferred to punch cards, checked for errors and sorted with respect to map unit, coordinates and serial number. A logical record in the data base will contain the information on one rock type from a locality.

All selections of information from the data base are handled by a program package called INTEREST. This system is characterized by an advanced retrieval facility in combination with a great number of statistical routines. The arguments for selections may consist of observation references, fixed arguments and conditions.

The retrieval programs of INTEREST have been completed with a number of routines for listing and plotting of geological information. Any information of the field sheet can be plotted with reference to the coordinates.

OBJECTS KATRINEHOLM SW

N CO-OR	E-CO-OR	ALT.CO-OR	OUTCR.NO	G	SPECIMEN	ANALYSIS CHEM MICR	PHOTO	SKETCH	DESC.	IMPORT. OUTCR.	OTHER
4620	145		157	S							X
5200	5		159	S			X	X		X	
5233	50		161,A	S	X	X		X			
			161,B	S				X			
6260	9660		162	X	X		X	X		X	X
6380	9480		167	S				X			
6420	9480		168	S			X	X			
6420	9420		169	S				X			
6500	9430	330	170	S	X					X	
6500	9500		171,A	S				X			
			171,B	S				X			
6008	9470		177	S							X
6080	9420		180	S			X	X			
6250	9110		187	S				X			
6175	8770		189	S				X			
8040	8045		192	S	X						
7690	8210		193,A	S	X						
8006	6570		194	S			X				
6150	8780		196	S	X						
6140	8720		198	S				X			
6130	8710		199,A	S				X			
			199,B	S				X			
6130	8670	478	200	S				X			
			200,B	S				X			
			200,C	S				X			
6520	8840		204,A	S				X			
			204,B	S				X			
9600	310	709	208	S	X						X
9705	9670		209	S	X						
9640	5770		210	S	X						
6840	7400		211	S	X						
6970	7750		212	S	X						
6950	7940		213	S	X						
6950	7980		214	S	X						
7695	8215		215,A	S	X			X			
			215,B	S	X			X			
7070	8725		216	S	X						
7065	9150		217	S	X						
6880	9030		252	S				X			
7690	9000	836	263	S			X				
7980	8905		266	S			X				
9500	7200		284,B	S	X						
7796	7800		307	S			X				

Figure 3. A standard cross list of objects.

Case Study

The potential of the GEOMAP system for displaying different results of analyses and retrieval is partly illustrated by the following output from the map unit Katrineholm SW, Central Sweden. The rocks of this area (25 km^2) are Svecofennian-age migmatites and various granites. The work was carried out during three field seasons and the number of observations is about 7 000.

At the end of the third field season the geologist in charge ordered some of the standard lists included in GEOMAP, for example, a comprehensive list of all decoded information and notes from each locality (Figure 2). The computer also produced a cross list of objects recorded on the field sheets (Figure 3) and separate lists of each object. The latter contained every note made with reference to the specific object. Moreover, the geologist obtained a list of all tectonic data including dykes. A number of equal-area stereo net analyses were made for linear and planar structures selected with respect to areas and rock types.

Standard routines were used for plotting maps of outcrops, observed rock types and tectonic data (Figure 4). The plots were made on transparent plastic to make it possible to superimpose them directly on the field map drawings. Together with the standard output the computer also plotted structure codes of migmatites together with type and quantity of mobilisate.

Some statistical routines of the INTEREST package were used for making cross-tabulations and histograms of migmatite structures, characters and quantities for different rock types in the area. Size and quantity of megagrain minerals were analyzed against habit for a number of granites. The geologist also got tables of frequency of colour and grain size for some rock types.

Experience with GEOMAP

The GEOMAP system has now been used during four field seasons by more than 50 geologists in Scandinavia. Besides field mapping, successful attempts have been made to introduce the system for educating geology students at universities in Sweden and Norway.

The present situation is that the geological surveys of Sweden and Norway are implementing the system for running on their own computers while the Survey of Finland is still testing the system. Boliden Mining Company is currently using GEOMAP for their prospecting and will soon take over a copy of the programs. Outside Scandinavia several organizations have expressed interest in GEOMAP and are now offered an English version.

Due to requests from the users in Scandinavia the field data sheet and some routines in the system were slightly modified in 1972. The important improvement is the division of the information with respect to observed rock types and the increase of tectonic registration. Another is the possibility for the geologist to use his own codes in some entries.

The field data sheet has been tested in various geological terrains of Svecofennian as well as Caledonian mountain rocks and seems to function well. Although the sheet was originally designed for igneous and metamorphic rocks, small complementaries were needed to adapt it for sedimentary rocks.

Among the users, a few geologists have maintained the opinion that the field sheet restricts the possibilities to make good observations. Most geologists, however, have found the combination of checklist and registration of notes sufficiently flexible for their work. For students, the input document has been a support in the field, which has been reflected in the amount and quality of their recorded data.

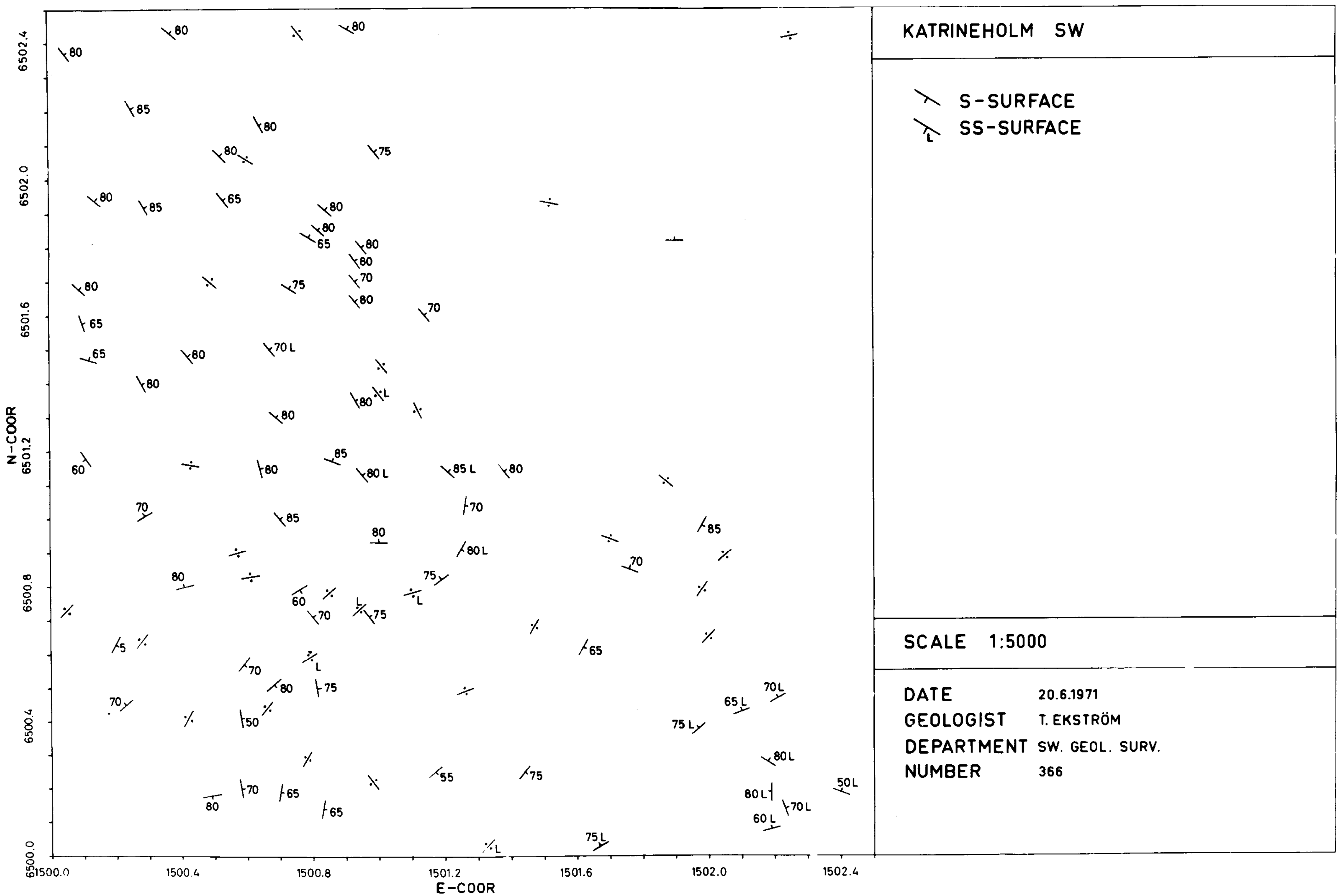


Figure 4. Computer-plotted map of planar structures from a test area.

Careful training is needed before geologists are able to make full use of the GEOMAP system. The amount of experience is obviously reflected in the number of errors and finally in the costs of processing of the collected material. Calculated on a collection of 500 outcrops, the cost of punching, storing and retrieval of a standard package of lists, plots and stereo net diagrams is about U.S. $0.75 per outcrop. Although the system is very user-oriented and easy to handle, the geologist usually prefers to order his jobs from a computer programmer. We are now working toward a closer collaboration between the geologist and the computer, using a time-sharing or interactive data system. This will reduce the costs and also train the geologist to use all the benefits of the computer system.

GEOMAP has a sister system called COREMAP. It is a data system for recording and processing information from drill cores and boreholes. The system may also be used in tunnels, adits, etc. The standard sheets include general, geological, chemical, joint and borehole deviation data but these may be extended to any other type of information. In COREMAP it is also possible to computerize notes.

GEOMAP Documentation: Publications

Berner, H., Ekström, T., Lilljequist, R., Stephansson, O., and Wikström, A.
1971 Data storage and processing in geological mapping. I: Field data sheet II: Data file. Geol. Före. Stockholm Förh., 93, p. 85–101 and p. 693–705.

Berner, H., Ekström, T., Lilljequist, R., Stephansson, O., Wikström, A.
1972 GEOMAP – A data system for geological mapping: Proc. 24th Inter. Geol. Congress, Section 16, p. 3-11, 6 figs.

Stephansson, O., Ekström, T., Berner, H.
1971 Computer techniques for geological mapping: Hornicka Program Ve Vede a Technice, Tjeckosloakia.

Stephansson, O., Ekström, T., Berner, H., Lundin, S.E.
1972 ADB-system för geodata: IVA Bergmekanikkommittén. Diskussionsmöte, Feb.

System Reports

Berner, H., Ekström, T., Lilljequist, R., Stephansson, O., Wikström, A.
GEOMAP: the user's guide to field data sheet.

University Data Centre: Geodata Group
GEOMAP Manual.

Ekström, T.K., Wirstam, A., and Larsson L.-E.
COREMAP – A data system for drill cores and bore-holes (in press).

Appendix

The system GEOMAP consists of two main parts, the first being a data modification and updating system, the second an analyzing and retrieval system.

Storing of the Recorded Field Data

The geological information recorded on field data sheets and later punched on cards is sorted into special units before it is fed into the computer. In Sweden the information is stored in units corresponding to topographic maps of the scale of 1:50 000. At the sorting, each unit is assigned information on name of map unit, name of geologist involved and year of mapping. The programs GESORT 1 and GESORT 2 are used for further processing and sorting. They also transfer the geological information to magnetic tape or disc pack. Error control is done by the program GUPDAT. The correct data are added to the data file while errors are listed for correction.

Analysis of the Stored Field Data

All geological data which are to be processed and presented in the forms of lists, plots, diagrams and statistical analyses are stored on magnetic tape. The processing is made by the INTEREST II system to which a number of special routines have been added.

With a few exceptions the routines of the GEOMAP system are coded in IBM System 360 FORTRAN IV (G). Memory required for the modification part is 158 K and for the calculating part 208 K.

An earlier version of the GEOMAP system is implemented on a CDC 3600 computer.

Implementation on IBM 370/155

Configuration:

1 card reader
1 card punch unit
1 line printer
1 plotter
2 magnetic tape devices
1 direct access device (or possibly more)

The system can with small reorganization use only discs, but it is preferable to use tapes for storage during longer periods.

The routines of the system use the following units (the logical number is given in parentheses).

GESORT 1

Input: Card reader or if card images are used disc or magnetic tape (1).
The new information (data).

Output: Disc (or magnetic tape).
The information is modified and sorted.

Scratch Units: Disc.
These units are normally controlled by the SORT-system.

GESORT 2

Input: Disc (or magnetic tape) (2).
Output from GESORT 1.

Output: Disc or magnetic tape.
The information is further modified and sorted.

Scratch units: Disc.
These units are normally controlled by the SORT-system.

GUPDAT

Input 1: Card reader (5).
Tables with certain keys to the codes from the field sheets (about 500 cards). The first card controls the logical unit number for INPUT 2, INPUT 3 and OUTPUT 1 below. The numbers in parentheses are not needed for these units.

Input 2: Disc or magnetic tape (1).
The output of GESORT 2.

Input 3: Magnetic tape (3).
The information to be updated, i.e. the old main file.

Output 1: Magnetic tape (2).
The new or updated information.

Output 2: Line printer (6).
For errors.

INTEREST

Input 1: Card reader (5).
Program control cards.

Input 2: Magnetic tape or disc (2).
The recorded data.

Input 3: Disc (11).
Tables for different codes used on the field data sheet.

Output 1: Line printer (6).
Lists, etc.

Output 2: Disc (42).
Pam-file of gefüge records.

Output 3: Magnetic tape (10).
The result for plotting.

Scratch units: Disc (e, 4, 8, 9, 12, 40, 41).

GEFÜGE

Input 1: Card reader (5).
Program control cards.

Input 2: Disc (42).
Pam file of gefüge records.

Output: Line printer (6).
Befüge diagrams.

Use of geological field data systems in Finland: experiences and problems

Fredrick Pipping
Geological Survey of Finland
SF - 02150 Otaniemi
FINLAND

Field Data Systems

Since 1968 the Finnish state-owned mining company Outokumpu Oy (base metals/sulphide ores) has been developing and designing an input, retrieval and processing system, GEOKU, for exploration mapping. It is now used by all the field parties of the company. An input document with a fair amount of data to be recorded is used. It is especially flexible with regard to the recording of tectonic and structural data. The system uses authorized lists of mnemonic codes that were established earlier for minerals and rocks and also for structural and textural terms. The storage program includes an automatic search for errors in the material, and rejecting storage before corrections have been made. The field data file is designed to save storage space, and as a result the data to be processed have to be retrieved through a special output file. The output is largely centred on the graphic presentation of the data. Structural data are displayed as maps and diagrams, petrographic and exploration data as symbol maps. The system is described in more detail in a recent publication by Gaal and Suokonautio (1973).

Another field data system, GEOULU, has been developed at the University of Oulu as a result of cooperation between the Geology and ADP departments. This project was financed by a special government grant for a survey of layered mafic intrusions in northern Finland. The GEOULU system is also largely conventional with a special input document, coding, etc. It seems likely that GEOULU and GEOKU will be combined at some future date. An agreement on cooperation and exchange of software has already been signed.

At the Geological Survey of Finland the department for bedrock mapping has participated in the Swedish GEOMAP system. Since 1971 GEOMAP has been tested in the field, and the Geological Survey has also shared some of the expenses for its development. However, a few practical difficulties have arisen, the most problematical of which are machine dependence and differences in hardware. GEOMAP was developed on Control Data and IBM machines, but the Geological Survey of Finland has a Hewlett-Packard unit. The data have to be processed in Sweden, which is highly impractical when experimenting on a large scale with a field data system! The situation will probably be improved by employing a programmer and a systems analyst in the mapping department, thus aiming at the full coordination of the computer processing of geoscience data at the Geological Survey.

A committee appointed by the Research Council of the Mining and Metallurgical Society of Finland to review the use of ADP in geological mapping in Finland should also be mentioned. In the committee report (dated February 1973) some technical recommendations on the input and storage of geological field data were made. However, the idea of a national system was rejected for the time being.

This is, in short, the present state of geological field data systems in Finland.

Present and Future Problems

A few remarks on the experiences and problems of using geological field data systems may be appropriate. These remarks are made from the point of view of a field geologist working in Precambrian terrain where less than 4 per cent of the bedrock is exposed.

1. It seems that the solution of how data are recorded on the outcrop is crucial for any geological field data system. Present techniques demand that the field geologist translate his concepts and terminology into codes or other forms of "computer language". It must be pointed out that the use of coding at this stage manifestly interferes with the very process of the geologist making and noting his observations. Geological language and vocabulary are stringent and purposeful, and generally international. A system that attempts to break down the entirety of accumulated scientific experience and definitions that constitute the geological vocabulary cannot work.
2. In geological field work, each region presents its special problems and demands an "individual" approach. This is true not only for the general physical conditions but especially so for the geological millieu. The problems and concepts of research in the Precambrian, Paleozoic, Mesozoic and Quaternary are similar but far from identical.
3. All geological field work, in whatever context it is performed, has a certain purpose. This purpose leaves its first clear marks, whether intended or not, on the recording of data. Geological field work is always project-oriented in one way or other, and geologists tend to be specialized, for example, in petrology, stratigraphy, structural geology, paleontology, etc. This necessarily leaves its mark on the field work. It should, moreover, be borne in mind that as both the purpose itself and the methods of achieving this purpose may change considerably in the course of the field work, flexibility of the system is imperative.

Points like these have to be considered when trying to design a geologic field data system that will work. A geologist's way of thinking in terms of interpolation and extrapolation with his data does not always agree with the logic of the computer, but one could hardly accuse the results of geological field work of being illogical. It should be remembered that a geological map is read in three physical dimensions, with a time dimension added, but that the informational units contained are counted by the dozen or more. Geologists tend to reshape methods and tools they borrow from other branches of science to suit geology and the earthy habits of geologists. The flexibility and capacity of computer techniques are certainly great enough to transform ADP into a useful tool for field geologists. We should soak the computer with the logic of geology, rather than computerize the geologist.

Reference

Gaal, G., and Suokonautio, V.
1973 An automatic data processing system for explorational mapping in Precambrian terrain: GEOKU. Geol. Surv. Finland Bull. 226, 26 p., 8 figs.

Selected geological field data systems in Canada: A brief description of their capabilities and objectives

W. D. McRitchie
Department of Mines, Resources and
Environmental Management
Geological Survey Section
993 Century Street
Winnipeg, Manitoba R3H 0W4
CANADA

A number of systems now exist in Canada for the machine-processable collection, storage and retrieval of geological field data.

The brief description of some of the systems presented here is largely based on presentations made at meetings of the National Advisory Committee on Research in the Geological Sciences (NACRGS)-appointed Working Group on Geological Field Data, and from replies received in response to a questionnaire that was distributed to the members of this working group.

The systems have been grouped for convenience of discussion on the basis of their geological objectives into those required for:

1. Regional Mapping Programmes
2. Detailed Stratigraphic Studies
3. Subsurface drill/well hole correlation
4. Mineral exploration, property examination and detailed mapping.

Regional Mapping Programmes

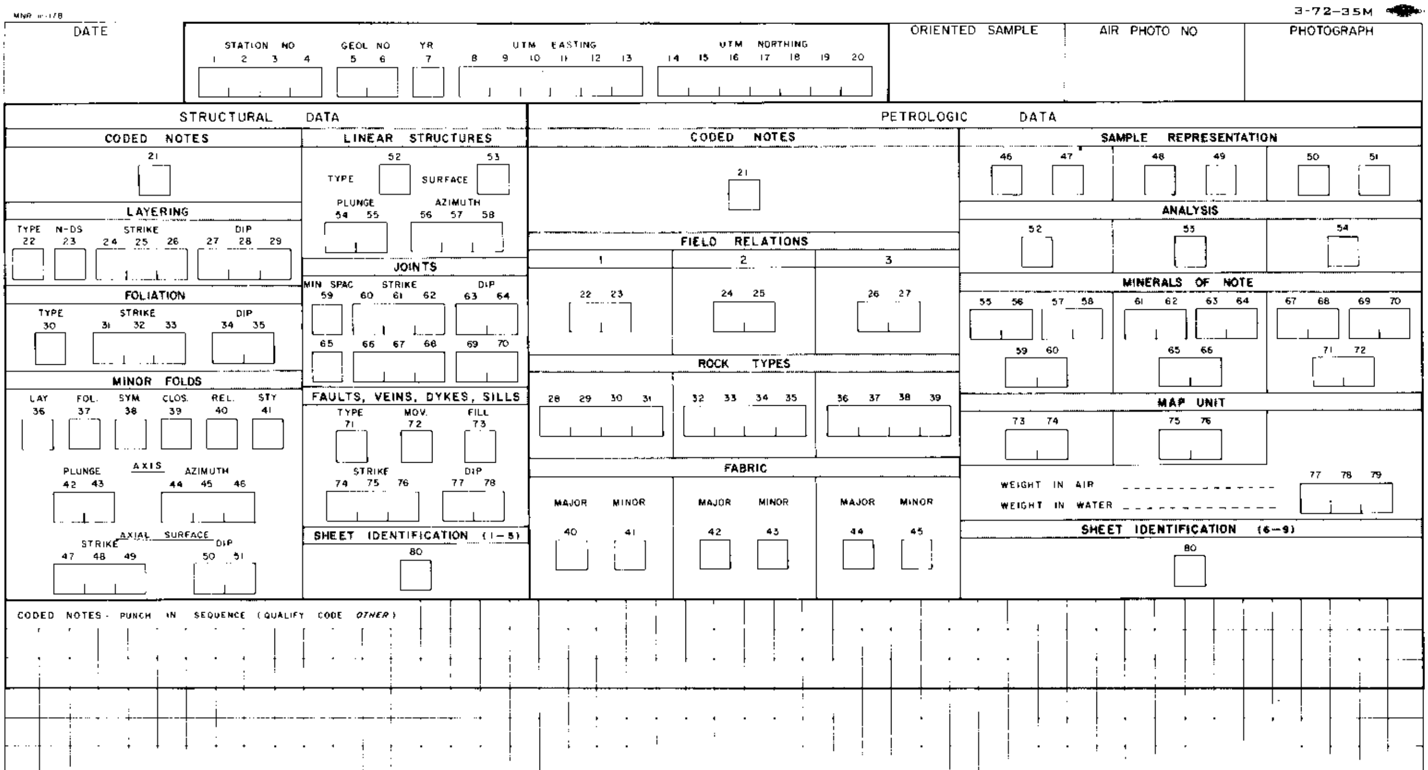
Some idea of the scope of these projects and the nature of their limitations can be gained from Table 1.

In nearly all such programmes definitive coverage is precluded by time and budget restrictions. The geologist must work with data collected from discrete points or stations augmented by limited continuous profiles obtained along the lines of the ground traverses. The quality and intensity of the data points varies from continuous traverses spaced at 1/2 km intervals to isolated stations 8 km apart.

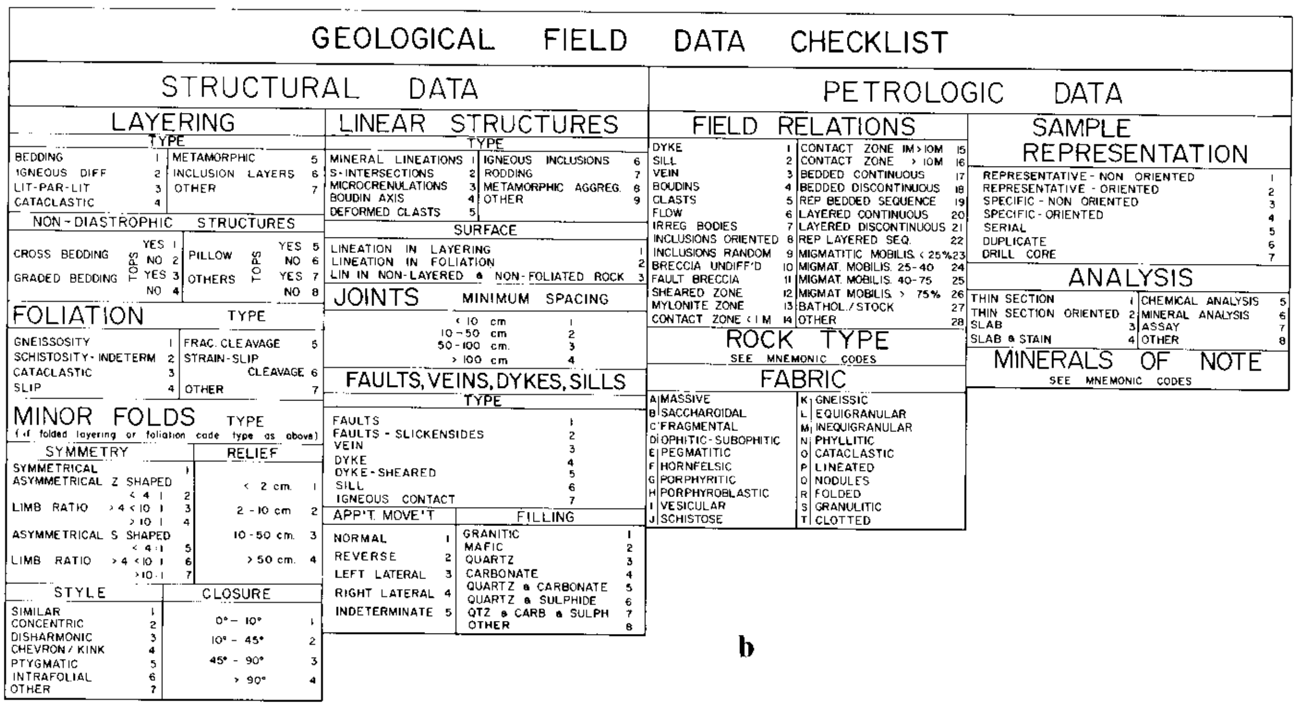
The task is such that the geologist is confronted, in an exceptionally reduced time spell, with considerable problems of standardization, consistency and data volume saturation.

Table 1. Summary of Canadian field data projects using computer techniques

Project/Organization	Area Covered	Term	No. of Geologists
Coast Mountains Project (GSC Vancouver)	43 000 sq. miles	10 yrs.	2-8/yr.
Grenville Project	12-15 000 sq. miles/yr.	6 yrs.	12/yr.
Superior Project	4 000 sq. miles	continuing	9/yr.
(Québec Dept. Nat. Res.)			
Project Pioneer	1 000 sq. miles	3 yrs.	12/yr.
Burntwood Project	7 500 sq. miles	3 yrs.	7/yr.
Southern Indian Lake	6 000 sq. miles	2 yrs.	20/yr.
(Manitoba Geol. Survey)			
Northern Appalachians (New Brunswick Min. Resource Branch)	?	continuing	24/yr.



a



b

Figure 1. Field document (a) and checklist (b) currently being used by the Manitoba Geological Survey.

In response to a common need for within-project standardization, a need to control and guide their assistants' data gathering, and a very great need to work rapidly with large volumes of newly collected data, a number of free and fixed format field data collection documents have been developed.

These documents are printed ahead of the field season and used in conjunction with check lists on which the various data parameters are coded. The current Manitoba format is illustrated (Figure 1) as an example of a typical document in which the data items are kept separate from a flipsheet checklist. Seasonal assistants are briefed in the use of the documents ahead of the field season and their working efficiency monitored during the early weeks of actual use. The documents are commonly verified either weekly or at the end of the season and are submitted for key-punching subject to scheduling with the respective computer facilities. Machine processing of the transferred data is conducted using both punched cards and tapes.

In Manitoba after the documents are returned from the computer centre they are bound into compact note books holding 200-250 documents. These books are used extensively thereafter by the project geologists and can be stored conveniently.

The Manitoba system is typical of many in that it makes possible, after initial data verification, edit and corrective procedures:

1. XY plots of structural, petrological or other parameters on selected coordinates at any scale;
2. A variety of statistical plots.

Additional programs have also been written or adopted from outside packages to handle input from geochemical field documents, chemical and modal analyses, and specific gravities. Output consists of a variety of normative calculations and other petrological parameters. At the present time the Manitoba system lacks both fully developed free formated note recall and search capability and a fully automated cartographic output designed for the production of final maps.

Detailed Stratigraphic Studies

Peter P. David (Université de Montréal) and Jacques Lebuis (Québec Department of Natural Resources) have designed a system for the standardized recording of complete stratigraphic information. Their system LEDA is exceptionally comprehensive yet functional, giving a level of data far beyond that possible in the regional mapping programmes. Localities recorded from the Quaternary deposits in the Ste. Anne des Monts region, Gaspé, in a single season are numbered in hundreds rather than thousands. The system, described as being user- and computer-oriented, comprises seven cards on which groups of related parameters have been compiled; e.g. card set 1, locality specifications; card set 2, unit descritions; card set 3, lithology descriptions and orientation, etc. The number of cards within any one set can be varied as required. The system can be used to cover nearly all unforeseen possibilities and weekly verification procedures are considered of prime importance.

Subsurface Drill/Well Hole Correlation

A dozen or more systems designed to such applications are now operational in Canada. Most are referred to in CODATA Bulletin 8(1972). Two additional systems include those of Mobil Oil Canada Ltd., and of the Geology Division of Alberta Research (formerly Research Council of Alberta).

Mobil's main concern lies with subsurface data from oil wells, but they have also used computerized techniques for mapping surface exposures in the Arctic. Mobil's data are merged with data from the file of Canadian Stratigraphic Services Ltd. which is pertinent to their interests in specific horizons. The most important techniques are those of facies analysis using a Calcomp flatbed plotter for the final output.

The Alberta EDMAT data file was established as part of an urban geological study of borehole and outcrop data in the Edmonton area. Rapid map production is now possible from stratigraphic borehole data coded in a fixed format.

Mineral Exploration, Property Examination, Detailed Mapping

The fully functional Rio Tinto (R.T.) System was established to handle data from all possible geological environments. The extreme flexibility of the free-format system is one of its most attractive facets. Within any one file there are a number of data RECORDS (stations), each of which contains a number of data SETS broken down into fixed sequences of PARAMETERS. Additional SETS can be sepcified as required and the PARAMETERS fields expanded or increased in number, depending on the data requirements. Retrievals can be made for a number of statistical and graphical manipulations. At present, Rio Tinto has about 30 key (SET) words which can be used to describe almost any geological observation. A thesaurus of allowable terms and within-project standardization processes ensure a maintenance of data compatability. Geochemical data are recorded in fixed-format but is changed to free-format for processing by the computer. Geologists working in areas with which they are already familiar preprint the data SET categories as checklists or guidelines to the type of information that they wish their assistants to record.

B.P. Minerals geologists in Montreal use a self-carboning standardized data collection document based on an original devised by the Institute of Geological Sciences in London. A number of variations on this basic format have been designed for specific activities. The system is designed to store and list field and analytical data and for creating a tape file, plotting and contouring and analyzing.

Geolog and Geosystem have been designed to handle both exploration and development projects, surface, open-pit and underground geological, geochemical and geophysical surveys in addition to the logging of drill core, cuttings and down-the-hole probes. The data are integrated with the topographic and geomorphological data of the project areas. Output consists of a variety of graphical or statistical displays and manipulations.

The Grenville System operated by the Québec Department of Natural Resources, originally developed at Queen's University, is one of the most successful field data systems. It has successfully processed large volumes of field data taken from the complex metamorphic Grenville province of southern Québec and enabled more efficient study by research and economic geologists.

In Ontario a system is being developed by the Ontario Division of Mines for application in the mapping of 1/4 mile to 1 mile map areas and presentation of yearly reports. The system is SAFRAS-based, and includes a double-sided input document. It represents a successful conclusion to efforts which were initially frustrated in trying to apply the early comprehensive fixed-format concepts, put forward by regional geologists to detailed property examination.

In Manitoba the procedures primarily developed for Project Pioneer and subsequently modified through several years of development have been found to be equally compatible with the detailed mapping of individual 15-minute map areas and to more regionally directed projects.

In a similar way the New Brunswick system has also met with considerable success in application to detailed mapping projects.

Concluding Remarks

In general the data collection documents have been modified repeatedly over the years since their initial inception and a number of relevant trends can be identified.

1. Standards are generally restricted to three simple conventions:

    1. Station numbers are right-justified numerics;
    2. Positions are recorded using Universal Transverse Mercator (UTM) coordinates: zone, easting and northing;
    3. The strikes of planar features are recorded as azimuths with dips to the right (including dips greater than 90o).

    These conventions greatly simplify the development of standard plotting programs to handle input from a variety of projects.

2. Decrease in emphasis in reliance of fixed formats except in application to quantitative data, location coordinates, etc.

3. Emphasis on ease of use in field; limited or no duplication of entries.

4. Separation of data into discrete fields for factual measurements, qualitative descriptions and sketches; this is based on the recognition that only specific categories of raw data reliably lend themselves to machine processing in bulk.

5. Emphasis on "in house" standardization as a means to making the data compatible within the sphere of the project rather than attempting to define universal standards. Documents are designed to specific projects.

By using these documents the survey geologist is now able to access and gain a feel for his data both manually and by automated means with much greater facility, but the main task of structuring his observations is facilitated whether he has access to a computer or not. This is a most important achievement.

The arguments in favour of machine-processable files more often than not revolve around the processing, archival and recall capabilities of the systems but it is equally important to stress the greatly enhanced qualitative and quantitative aspects of the data gathering phase and the ease with which the improved organization of the raw data lends itself to a hitherto unparalleled opportunity for checking and validating the raw information prior to subsequent processing.

The full potential of many completed field data files is still limited, however, by inconsistencies in the more descriptive data items. Classifications evolve throughout the field season and too often records of when "tonalite" was renamed "diorite" or when that funny pink mineral was finally identified as andalusite, are not updated or even committed to writing. There is therefore a fundamental need to remind field geologists to record every instance when, as a result of new observations, their ongoing classifications are modified. The updating of the existing file can then be accomplished manually or by machine, but most important of all it can be updated or merged reliably.

Table 2. Characteristics of Canadian field-data systems.

Organization	City	Respondents	System	Theme	Purpose	Physical Support
Geological Survey of Canada	Vancouver, British Columbia	J.A. Roddick W.W. Hutchison	Coast Mountain project, field data on an igneous-metamorphic complex	carto/petrog	Production Data analysis Archival	UNIVAC 1108 and IBM 360/68 under M.T.S.
Québec Dept. of Natural Resources	Québec, Québec	K.N.M. Sharma A.F. Laurin H.R. Wynne-Edwards	Computerized geological mapping in the Grenville Province, Québec	carto/struct	Data analysis Production Archival	IBM 360/50 UNIVAC
New Brunswick Dept. of Natural Resources	Sussex, New Brunswick	A.A. Ruitenberg	Caledonia mountain/Bathurst; field data igneous and metamorphic complexes	carto/struct petrology	Archival Data analysis Production	IBM 360/50
Manitoba Dept. Mines Res. & Environmental Management - Geology Division	Winnipeg, Manitoba	W.D. McRitchie I. Haugh J.S. Roper	Project Pioneer, Burntwood Project Southern Indian Lake field data on Precambrian metamorphic complexes	struct/carto petrology geochem.	Data analysis Data rectification Archival	IBM 370/155 370/135 operating on OS
Université de Montréal	Montréal, Québec	P.P. David J. Lebuis	LEDA field data from Quaternary deposits in Gaspé Peninsula	carto/sedim.	Data analysis Production Archival	one of flux
Mobil Oil Canada, Ltd.	Calgary, Alberta	D.A. Clark	Correlation of sedimentary borehole data in defining target horizons	borehole/oil	Production Data analysis Archival	IBM 360/50 2250 Graphics 718 Flatbed
Research Council of Alberta - Geology	Edmonton, Alberta	J.W. Kramers	Correlation of borehole and selected outcrop data for environmental study	borehole/outcrop correlation	Production Archival	IBM 360/67 under M.T.S. Calcomp drum plotter
Rio Tinto Canadian Exploration Ltd.	Vancouver, British Columbia	R.V. Longe R. Willie R. Hewton	Field observation data	min. res./carto	Data analysis Archival	UNIVAC 1108 ASR 33 terminal
B.P. Minerals Ltd. B.P. Oil	Montréal, Québec	D.M. Taylor	Field and analytical data	min. expl. geochron.	Archival	IBM ?
Canada Geosystems Ltd.	Langley, British Columbia	P.H. Blanchet C.I. Godwin	Exploration and development data; mineral deposits	min. expl.	Data analysis	?
Ontario Division of Mines	Toronto, Ontario	N. Trowell	Field observation data	struct/carto petrography	Production Archival	UNIVAC 1106

Bibliography

Ambach, H.
1972 Description of computer programmes: Man. Mines Br., Unpub. Rept.

Grice, R. H.
1971 Geological data handling in urban areas: Canadian Geotech. Jour., v. 8, no. 1, p. 134-138, 2 figs.

Haugh, I., Brisbin, W. C., and Turek, A.
1967 A computer-oriented field sheet for structural data: Canadian Jour. Earth Sciences, v. 4, no. 4, p. 657-662, 2 figs.

Laurin, A. F., Sharma, K. N. M., Wynne-Edwards, H. R., and Franconi, A.
1972 Application of data processing techniques in the Grenville province, Québec, Canada: Proc. 24th Inter. Geol. Congress, Section 16, p. 22-35, 9 figs.

Lebuis, J., and Davis, P.
1973 Systeme LEDA: Utilisation des fiches standardisées pour enregistrer les données géologiques des dépôts quaternaire: Ministerre des Richesses Nat. du Québec Technical Document.

McRitchie, W. D., Turek, A., Haugh, I., and Brisbin, W. C.
1970 Recording petrologic and structural data for machine processing; Project Pioneer, Manitoba Mines Br., unpublished report.

Roddick, J. A., and Hutchison, W. W.
1972 A computer-based system for geological field data on the Coast Mountains, British Columbia, Canada: Proc. 24th Inter. Geol. Congress, Section 16, p. 36-46, 5 figs.

Sharma, K. N. M., Laurin, A. F., Wynne-Edwards, H. R.
1973 Computer-based geological field data system for the Grenville Project, Quebec: an operations manual; Ministerre des Richesses Nat. du Québec Publ.

Wynne-Edwards, H. R., Laurin, A. F., Sharma, K. N. M., Nandi, A., A., Kehlenbreck, M. M., and Franconi, A.
1970 Computerized geological mapping in the Grenville Province Quebec: Canadian Jour. Earth Sci., v. 7, no. 6, p. 1357-1373, 10 figs.

Computer-based field data files in the Regional and Economic Geology Division, Geological Survey of Canada

T. M. Gordon
Geological Survey of Canada
601 Booth Street
Ottawa, Ontario K1A 0E8
CANADA

The Task

Two years ago a study was initiated in Ottawa to investigate the use of computer technology in expediting the collection, storage and retrieval of geological field data for the preparation of geological maps and reports. This paper will outline the approach taken to this project in the context of the Geological Survey's goals and constraints.

The Goals

The Regional and Economic Geology Division of the Geological Survey of Canada has as one of its tasks the presentation of geological information on 1:250 000 scale maps. Much of the work based in Ottawa involves mapping in Precambrian terrain – gneisses, greenstones, metasediments and intrusive igneous rocks. An average map sheet has an area of 11 000 km^2 with field work normally taking three to six months to complete. Helicopter traversing is used extensively, resulting in observations with an average spacing of 3 km. A single geologist is responsible for compiling a map and report, although he will have two or three assistants also collecting field data. The primary users of the geological maps are mining and petroleum exploration companies, as well as resource potential evaluation groups in government, both of whom require syntheses and interpretations of geologic data rather than primary field observations. This, combined with the reconnaissance nature of the work, minimizes the requirements for a comprehensive file of raw field data.

The Constraints

Geological. The wide variety of geological provinces in the Canadian Shield, each with differing lithologies and tectonic history, makes it impossible to define a standard mapping technique. Because of the limited time available for field observations, the raw data to be collected must be chosen to optimize the information that can be derived from them.

Psychological. Historically, Canadian field geologists have been strong individualists – as might be expected in a country where the geologist was often the first scientific observer in unexplored wilderness. This attribute has been sustained to the present by the "project approach" to geological mapping which requires a single scientist to be responsible for all phases of a mapping project – from logistic arrangements to scientific interpretation. The production of a map is thus a highly personal task – with no one more qualified to make geological decisions about the project than the scientist in charge. There is thus no practical way to arbitrarily impose standards for the recording of field data on this group of field officers.

Hardware. The Department of Energy, Mines and Resources, (EMR), of which the Geological Survey of Canada forms a branch, operates its own Control Data Corporation Cyber 74 computer. It is departmental policy that virtually all computer work be done on this "in-house" installation. This prevents use of some important commercially available data-base management systems which are specific to other manufacturers' equipment.

Software. The Regional and Economic Geology Division does not have its own programming staff, but instead competes with other EMR users for programming services provided by the departmental Computer Science Centre. Although the quality of programming is good, resources are scarce, hence there is often a time lag in program development, as well as a problem in setting priorities for various tasks.

The Computer Science Centre has installed and is knowledgeable with several generalized data handling systems, including MARS VI and SAFRAS.

The Approach

Over the past eighteen months, a series of pilot projects has been underway to assess the use of computer technology in expediting field mapping. From the first it was obvious that the wide variety of geological problems and mapping techniques precluded the use of any "standardized" input document or file structure. Each project leader designed his own file and system to suit his particular needs. In this way we hoped to "mimic" what the field men were already doing and avoid the pitfalls of grandiose super-system planning. This also gave us an extremely flexible mode of operation with virtually no geological constraints.

The price to be paid for this flexibility was additional effort by each individual in learning some fundamentals of computer programming. This required an educational programme designed specfically for field geologists. Fortunately, we have an excellent training officer in EMR.

Two sessions of an introductory FORTRAN course were given, attended by approximately 30 staff members. The course emphasized techniques for handling character-string information and input and output – essentials to any file-handling system. A total of five days, spread over two weeks, was required for the course. At the completion of the course, the students were capable of writing programs to edit, retrieve, and list data from their files. At the very least, they had become aware of the powers and limitations of computers and programmers.

Although there are several "generalized data-base management" systems (including SAFRAS and MARS VI) available to the GSC, it was decided that the best initial approach to field data files is to construct a card-image "fixed-format" file. Because field data are normally collected on preprinted checklists, the presence or absence of "free-format" capability is irrelevant. Highly structured card-image files have the advantage of being upward-compatible with any future developments in data handling. We have, in fact, developed methods for transferring these files to the MARS VI system and using the data validation and retrieval capabilities of MARS.

This approach is resulting in a convergence of file designs and some de facto "standardization" of file structure. Most files contain a maximum of three distinct card formats, corresponding to lithological, structural, and free-text data. A single field in all cards contains the station number, while a separate field identifies the format type and, if necessary, the sequence number of that particular card. Such a file is easily converted to fixed-length records on tape or disc using in-core read and write sub-routines.

"Standards" in the data itself are confined to three simple conventions. Station numbers are right-justified numerics; positions are recorded by UTM zone, easting, and northing; and the strikes of planar features are recorded as azimuths with dips to the right. These conventions greatly simplified the development of a standard plotting program to handle input from a variety of projects.

Work is continuing to develop generalized utility programs to provide field officers with useful report and graphic output from their data files. We hope that by involving individual scientists at all stages of development we will attain a useful and practical "system" for Geological Survey of Canada mapping projects.

Sub-Area Retrieval System (SARS) used on the Coast Mountains Project of the Geological Survey of Canada

W. W. Hutchison and J. A. Roddick
Geological Survey of Canada
100 West Pender Street
Vancouver, British Columbia V6B 1R8
CANADA

In late 1971 the Geological Survey of Canada initiated a programme of pilot projects to evaluate the usefulness of computer-based techniques for collection, storage and manipulation of geologic field data and information. The following is a description of the most highly evolved of these projects.

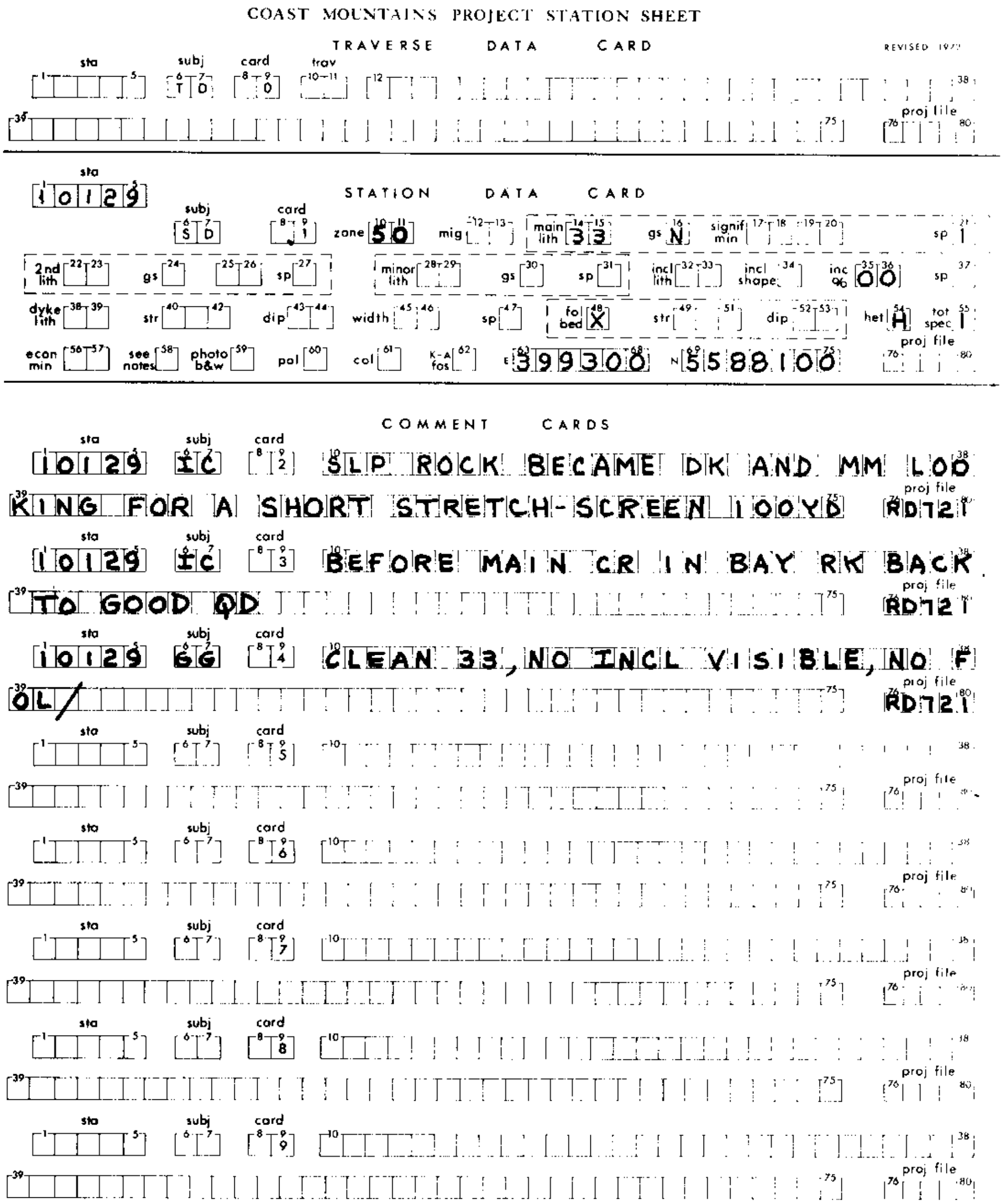
The system used in the Coast Mountains Project had been in existence since 1965, (Hutchison and Roddick, 1968). Major evolutionary stages in this system have been:

- 1965 - Computerization of limited amount of field data
- 1969 - Plotting of strikes and dips on stable base to go directly to the printer
- 1970 - Use of free text to complement data (all data and handwritten notes were keypunched at a cost of 2 per cent of field budget)
- 1973 - Compilation (using interactive terminal) by instantly retrieving the complete data base for a map-unit sub area at one command and by interrogation, data reduction and synthesis of the generated sub-file.

For the field geologist the last two stages of evolution were the most beneficial. Because these are potentially important and fundamental and because they are commonly lacking in systems developed to date, this report illustrates these aspects as an example of one pilot project of the Geological Survey of Canada.

Coast Mountains Project

The Coast Mountains Project of the Geological Survey of Canada was set up in 1962 to conduct geological reconnaissance mapping of the Coast Mountains in British Columbia from latitude 56° N to Vancouver (near 49° N). The area comprises about 43 000 square miles (110 000 sq. km.). So far data and information have been collected for 25 000 stations and 25 000 specimens. This information pertains to possibly the largest post-Precambrian plutonic complex in the world so


COAST MOUNTAINS PROJECT STATION SHEET

TRAVERSE DATA CARD

REVISED 197?

sta | subj T D | card 0 | trav | proj file

STATION DATA CARD

sta 10129 | subj S D | card 1 | zone 50 | mig | main lith 33 | gs N | signif min | sp 1

2nd lith | gs | sp | minor lith | gs | sp | incl lith | incl shape | inc % 00 | sp

dyke lith | str | dip | width | sp | fol bed X | str | dip | het H | tot spec 1

econ min | see notes | photo b&w | pol | col | K-A fos | E 399300 | N 5588100 | proj file

COMMENT CARDS

10129 IC 2 SLP ROCK BECAME DK AND MM LOO
KING FOR A SHORT STRETCH-SCREEN 100YD RD721

10129 IC 3 BEFORE MAIN CR IN BAY RK BACK
TO GOOD QD RD721

10129 GG 4 CLEAN 33, NO INCL VISIBLE, NO F
OL/ RD721

Figure 1. Typical input document for geological field data. This sheet corresponds to lines with asterices in Figure 2. Note, however, this document does not contain petrographic data which was merged and added into field data file as shown in Figure 2.

it was decided some effort should be made to systematically synthesize the vast quantities of data and information and maintain these for further study.

Evolution of System

The development of the system is documented in Roddick and Hutchison (1972).

Current System

Input. The present system using pre-printed check sheets using standard 80 column cards is still in use (see Roddick and Hutchison, 1972, for codes). An example of the current field station input document (Figure 1) is shown to illustrate the much greater allowance now made for recording free text. Station locations are measured at base camp using Romer measuring grids on a map printed with a Universal Transverse Mercator grid. A separate input document is used for petrographic information such as rock specimen data, specific gravity, mineral abundances and revised rock names (Roddick and Hutchison, 1972, Fig. 2).

Editing. Editing is the most time-consuming part of this endeavour. Prime information can usually only be checked by the field geologist and this results in his being severely handicapped. At present we do not have a wholly systematic approach to editing.

The three aspects of editing are as follows: editing of station locations, editing for invalid codes or numbers and scanning of some prime data and comments.

Editing can also be done during retrieval and this is described below.

Retrieval. For economical access to our now large master file, a system of on-line retrieval, called SARS (Sub-Area Retrieval System) was developed to run on the University of British Columbia's IBM 360/67 under the MTS (Michigan Terminal System) operating system. SARS, however, could be adapted to other comparable operating systems, as the techniques developed in its implementation are applicable to more general systems.

SARS consists of a master data file, several ancillary files, the on-line retrieval program, and file maintenance programs. It includes various features of the MTS operating system, particularly the file-handling capabilities and the MTS editor.

The master file represents a merge of both the station file and the petrographic file. It is a sequential file with line numbers related to station numbers. Previously developed programs permit us to plot most of the hard data, such as attitudes, dykes, mineral occurrences, etc., on either a drum or flat-bed plotter (the latter being used for direct plots onto a topographical base map).

To manipulate the information in the master file and to retrieve selected parts of it, a program called SUBMON was developed. This program performs operations on the master file consulting, where required, a file called 'Subspecs', which contains a list of stations pertaining to a lithologic unit. This file must be previously built by the user who selects and inputs the station numbers. This onerous task is somewhat simplified by a provision which makes ranges of consecutive stations acceptable, and could be done entirely by the computer in the type of geological work where the lithologic unit is known at the time the stations are made.

```
* 74   39-58  TOBA QD-GD
  10022SD0150  04CSP  1                        F33090 1       3987005585700RD701P12690804078103
  10022GG 2 ALMOST AUGEN STREAKINESS IN PLACES. ONCE EUHED BIOT HAS GONE TO F
  10022GG 3 INE GRAINED MAFIC PSEU DOMORPHICALLY
  10023SD0150  05F    1          00            X      SI      3976005582000RD701P12630213147104
  10023GG 2 VERY WHITISH LEUCOCRATIC ROCK/SUGARY TEXTURE
  10275SD0150  05F    1                        M33590S1       3977005581000RD701P126302151C7304
  10275GG 2 QM BUT MAYBE INSERTED BETWEEN QD AND GD/QUITE A FEW SCHLIEREN
  10274SD0150  33N    1          E02           G31060M1       3953805580450RD701P12771703057543
  10274GG 2 DEFINITELY NOT THE GD
  30292SD0156  33C    1          AME05         G33090 1       3856605588240RD701P12711201226543
  30292GG 2 STRONGLY FOL 33 -BUT NOT ALTERED
  30293SD0156  33M    1FP   2         CV03090  3G29090S3   2  3879205590450RD701P12690912166344
  30293SD0156  33M    1FP   2         CV03090  3G29C90S3   2  3879205590450RD701D3291        BA
  30293SD0156  33M    1FP   2         CV03090  3G29090S3   2  3879205590450RD701P22660905C97724
  30293GG 2 WELL FOL 33/ MINOR SCHLIEREN/
  30293DY 3 SWARM OF DARK GREEN DY- SYNPLUTONIC W CHILLED MARGINS-SEE
  30293DY 4 PHOTOS/1/3 OC IS FP POSS EQUIV OF PO GRANOD TO W THIS CUTS
  30293DY 5 THE 33 AND SP DY/
  30293PH 6 3-2-25A,26A DIKE W CHILLED MARGINS AND SYNPLUTONIC ÓFFSETS
  30395SD0156  24M    1               00 N 29090  2X     H2   3898005591310RD701P12680912097044
  30395SD0156  24M    1               00 N 29090  2X     H2   3898005591310RD701D2289        AN
  30395GG 2 CLEA HOMO 24- V MINOR AP VEINS/ALSO V MINOR EP STRINGERS/
  30395GG 3 DY SWARM WELL EX ON CIRQUE WALL- SOME AS WIDE AS 25FT-ALSO AL
  30394SD0156  24M    1          AMR01         I29090H1       3926805590540RD701P12640615186134
  30394GG 2 CLEAN 24-BUT TX DOES NOT LOOK NORMAL-MAFICS- ANHEDRAL
* 210129SD0150  33N    1          00           X      H1     3993005588100RD721P12741503156733
* 210129IC02SLP ROCK BECAME DK AND MM LOOKING FOR A SHORT STRETCH SCREEN 100YD
* 210129IC03BEFORE MAIN CR IN BAY RK BACK TO GOOD QD
* 210129GG04CLEAN 33,NO INCL VISIBLE, NO FOL/
  210130SD0150  33N    1                       E34590 1      3973505587650RD721P12691512155844
  210130IC02PASSED SUDDENLY FROM MASSIVE QD INTO EXTREMELY ELONG AGMATITE SCHL
  210130IC03IEREN GNEISS AND SCREEN GNEI SS/ E4 HAS INCL UP TO 10FT OR MORE IN
  210130IC04LENGTH
  210130GGC5ATP STRONGLY FOL TO GNEISSIC QD/ COARSE PATCHES OF FG MAFICS
  210130SU06THE TRAV SEEMED TO PASS DOWN SECTION THRU A THICK SEQUENCE OF ISO
  210130SU07CLINALLY FOLDED BI HO QTZ SC AND MINOR QTZITE WITH A COUPLE OF ZON
  210130SUC8ES OF GRANITOID RX/ SOME OF THE METASEDS ARE LIMY,SOME LAYERS DIOP
  210130SU09THAT SEQUENCE IS UNDERLAIN BY ABOUT 4000FT OF CONGL THAT CONSISTS
  210130SU12LARGELY OF QD OR GD CLASTS UP TO 3FT ACROSS IN A LT GREY SCHISTOSE
  210130SU13ARKOSIC MATRIX IN WHICH BIOT AND SMALL GARNETS HAVE GROWN/ THE CLA
  210130SU14STS ARE SOMEWHAT STRETCHED BUT HAVE NOT DEV ANY INTERNAL FOL/STRON
  210130SU15G FOL IN MATRIX BENDS AROUND CLASTS/ WE PASSED NEXT INTO QD THAT
  210130SU16MATCHES THE CLASTS,SUGGESTING AN UNCONFORMITY/ BETW 129 AND 130 IS
  210130SU17A 4 5C00FT WIDE ZCNE OF SPECTACULAR EL AGM AND SCHLIEREN GNEISS
  210131TD0016JUNE 8/ TOBA INLET UNCONFORMITY HUNT " CONT OF TR 15/ R.T./ CAVU
  210131SD0150  33N    1          AMR01         FC0C90S1      3997505588200RD721P12782002C96933
  210131AP0292K 40 81
  210131IC03FROM GOOD CONGL NE OF STA 128 TO HERE WE HAVE EXAMINED THE OTCP
  210131IC04CLOSELY,WALKING PART OF IT/ CONG SEEMS TO EXTEND TO THIS SIDE OF
  210131IC05128,BUT IT BECOMES VERY DEFFICULT TO RECOGNIZE BECAUSE MATRIX AND
  210131IC06CLASTS ARE OF SAME MATERIAL, AND ONLY ON CERTAIN W SURFACES CAN
  210131IC07CLASTS BE DISCERNED/ WOULD TEND TO PUT CONTACT IN NO OTCP AREA BET
  210131IC08W TP AND CREEK MOUTH TO NE
  210131GG09ATP CLEAN QD,HIGH QTZ AS IN CLASTS/ V FAINT FOL/ A FEW NEB INCLU/
  210132SD0150  43N    1          AME03         E34090H1      3967005587150RD721P12703C15134244
  210132AP0292K 39 165
  210132IC03FROM STA 130 TO HERE, SOS WITH SLIVER AM INCLUS/
  210132GG04ATP SOS,BUT BIOT DEF DOMINANT V NARROW SLIVER INCLUSIONS/ QD IS
```

Figure 2. To illustrate retrieval of complete data base for one sub-area, based on one command, this is start of listing for sub-area 74 – TOBA QUARTZ DIORITE – GRANODIORITE. Lines with asterisk appear in Figure 1.

```
:mts r nea
LINE DELETED
_mts
#r mean 5=-d 6=-74
#EXECUTION BEGINS
#EXECUTION TERMINATED
#1 -74
```

	STN NO	INCL	FOL	HET	SPGR	MAF	KFP	QTZ	ROCK
> 1	STN NO	INCL	FOL	HET	SPGR	MAF	KFP	QTZ	ROCK
> 2									
> 3	10022	0	F		2.69	8.	4.	7.	03
> 4	10023	0	X	S	2.63	2.	13.	14.	04
> 5	10274	2	G	M	2.77	17.	3.	5.	43
> 6	10275	0	M	S	2.63	2.	15.	10.	04
> 7	30292	5	G		2.71	12.	1.	22.	43
> 8	30293	0	G	S	2.69	9.	12.	16.	44
> 9	30293	0	G	S	2.66	9.	5.	9.	24
> 10	30394	1	I	H	2.64	6.	15.	18.	34
> 11	30395	0	X	H	2.68	9.	12.	9.	44
> 12	210129	0	X	H	2.74	15.	3.	15.	33
> 13	210130	0	E		2.69	15.	12.	15.	44
> 14	210131	1	F	S	2.78	20.	2.	9.	33
> 15	210132	3	E	H	2.70	30.	15.	13.	44
> 16	210133	0	M	S	2.70	18.	5.	10.	33
> 17	230206	2	E	S	2.66	12.	14.	21.	44
> 18	230207	3	E	S	2.71	13.	3.	6.	43
> 19	230208	0	F	H	2.74	15.	12.	19.	24
> 20	230209	1	G	S	2.68	21.	17.	12.	24
> 21	230210	0	E	S	2.65	7.	13.	12.	44
> 22	230212	0	M	S	2.69	9.	3.	17.	43
> 23	230213	5	F	M	2.81	21.	0.	2.	2
> 24	230214	1	X		2.84	22.	0.	2.	32
> 25	250133	0	M	M	2.66	10.	10.	12.	54
> 26	250134	20		M	2.68	15.	15.	12.	44
> 27	250135	8	F	M	2.64	10.	20.	12.	54
> 28	250136	5	M	M	2.68	15.	0.	10.	53
> 29									
> 30	MEAN OF	26 STATIONS			2.70	13.	9.	12.	

```
#END OF FILE
#(
LINE DELETED
```

Figure 3. This is a tabular summary of main features of this sub-area. (Abbreviations: STN No = station number, INCL = inclusions in percent of outcrop, HET = outcrop heterogenite – H = homogeneous, S = *slightly* heterogeneous, M = moderately heterogeneous, SPGR = specific gravity, MAF = total mafic minerals, KFP = potash feldspar, QTZ = quartz, ROCK = rock name – first digit is hornblende/biotite ratio, second digit is 3 = quartz diorite, 4 = granodiorite.

```
LINE DELETED
zed -s74DY
Tcol 8
Tm@a@nl /file 'dy'
:  30293DY 3 SWARM OF DARK GREEN DY- SYNPLUTONIC W CHILLED MARGINS-SEE
:  30293DY 4 PHOTOS/1/3 OC IS FP POSS EQUIV OF PO GRANOD'TO W THIS CUTS
:  30293DY 5 THE 33 AND SP DY/
: 230207DY05DARK GREEN SC DYKES FORM 3( SCREENS/ LOCALLY THEY ARE CROSS
: 230207DY06CUTTING
: 230209DY04NOTE DYKES ARE NOT SYN PLUTONIC BUT ARE PREMETA.            }
: 230209DY06AMPHIBOLITE DYKE SWARMS ARE COMMON THEY GENERALLY HAVE A     }  One
: 230209DY07CONSISTENT TREND BUT COMMONLY CROSS CUT THE QD LOCALLY THEY  }  station
: 230209DY08FORM A STOCKWORK BUT NOTE THE DYKES AND THE PLUTONIC ROCK HAVE }
: 230209DY09BOTH BEEN META. THIS PLUTON IS PROB PRECONGL                }
: 230210DY08DARK GREEN AM DY ARE COMMON THEY GENERALLY FORM SWARMS AND
: 230210DY09MAKE UP 10( OF OTCP
: 230212DY07TWO AGES OF DY PARTIALLY GRANITIZED DY THEN YOUNGER BUT META
: 230212DY08AN DYKE SWARMS SUCH AS AMPHIBOLITE IN SPECD2/ THESE DYKES DO
: 230212DY09NOT HAVE CHILLED MARGINS INSTEAD FINE GRAINED SCHISTOSE MARGNS
: 250133DY05ANDESITE THAT NOW HAS A'FOLIATION RESEMBLING AN AMPHBOLITE
: 250134DY05S FRACTURES ALONG SIDE OF DYS/ THEY IN THICKNESS FROM 6 INCHES TO
: 250134DY065 FEET/ THEY WEATHER TO A DARK GRAY%SOOTY COLOR</CHILL MARGINS
: 250135DY06ANDESITE/ SLIGHTLY PO WITH MUCH EPIDOTE " CHL/ SMALL PEG VEINS WI
: 250135DY07TH FRAGMENTS OF DIKE ROCK ARE A LATER CROSSCUTTING FEATURE
: 250136DY04BLACKISH GREEN WEATHERING/ MUCH EPIDOTE " CL/ MAKES UP 40( OF OT
: 250136DY05CP HERE/ TRENDS WERE SAME AS PREVIOUS STA/
:)
LINE DELETED
```

Figure 4. Retrieval of comments on dykes through searching two-letter subject code DY in columns 8 + 9.

```
LINE DELETED
zm@a@nl /file 'su'
: 210130SU06THE TRAV SEEMED TO PASS DOWN SECTION THRU A THICK SEQUENCE OF ISO
: 210130SU07CLINALLY FOLDED BIHOQTZ SC AND MINOR QTZITE WITH A COUPLE OF ZON
: 210130SU08ES OF GRANITOID RX/ SOME OF THE METASEDS ARE LIMY,SOME LAYERS DIOP
: 210130SU09THAT SEQUENCE IS UNDERLAIN BY ABOUT 4000FT OF CONGL THAT CONSISTS
: 210130SU12LARGELY OF QD OR GD CLASTS UP TO 3FT ACROSS IN A LT GREY SCHISTOSE
: 210130SU13ARKOSIC MATRIX IN WHICH BIOT AND SMALL GARNETS HAVE GROWN/ THE CLA
: 210130SU14STS ARE SOMEWHAT STRETCHED BUT HAVE NOT DEV ANY INTERNAL FOL/STRON
: 210130SU15G FOL IN MATRIX BENDS AROUND CLASTS/ WE PASSED NEXT INTO QD THAT
: 210130SU16MATCHES THE CLASTS,SUGGESTING AN UNCONFORMITY/ BETW 129 AND 130 IS
: 210130SU17A 45000FT WIDE ZONE OF SPECTACULAR EL AGM AND SCHLIEREN GNEISS
: 230208SU12AGMATITE " DIORITE 123 ON MIKES TRAV/ 124126 PALE PINK
: 230208SU13APLITIC GD/ 127132 DARK MAFIC PHASES DIOR/QD VARIED AGMAT
: 230208SU14THEN SHEARED V STRONGLY SHEAR FOL GRANOD THEN V SMEARED
: 230208SU15ALMOST LIKE ARKOSIC MATRIX
: 230208SU16NOTE  JIMS GRANOD 128 IS EDNETICAL TO CLASTS %EG. 30200E1<
: 230208SU17FOUND IN CONGL./ NOTE ALSO JIMS PRECONGL GD IS V SIM TO 30207
: 230208SU18%IE HIS 3 SAMPLES ARE SIM TO THIS SECTION<
: 250136SU02LL FOLIATED IS FOUND/ THEN TO THE END WAS A SHEARED PINK AUGEN OR
: 250136SU03THOGNEISS %GRANOD</ ALSO SHEARING OF ANDESITE DIKES IS FOUND/ NOTEE
: 250136SU04D FEATURES WERE THE ELONGATE AGMATITE ZONES AT STA 121 " 132 " TH
: 250136SU05E DIKE SWARM AT STA 134135
: 250136SU07THE TRAVERSE BEGAN AT SNOUT PT WITH A SLIGHTLY FOLIATED QTZ DIORI
: 250136SU08TE/ THIS CONTINUED FOR 3 STATIONS/ NEXT STATIONS WAS A FINE GR PI
: 250136SU09NK APLITIC FOLIATED GRANOD/ UNTIL STA 133 QTZ DIORITE DIORITE WEL
:
```

Figure 5. Retrieval of traverse summaries through searching two-letter subject code SU in columns 8 + 9. Note that first 10 lines could be almost read by someone not familiar with the system.

```
 mts
LINE DELETED
zmts
#ed -s74
:s@a@nl /file 'congl'
: 210130SU09THAT SEQUENCE IS UNDERLAIN BY ABOUT 4000FT OF CONGL THAT CONSISTS
: 210131IC03FROM GOOD CONGL NE OF STA 128 TO HERE WE HAVE EXAMINED THE OTCP
: 210133GG06DIST LARGE PATCHES/ RESEMBLES CLASTS IN CONGL/
: 230208SU17FOUND IN CONGL./ NOTE ALSO JIMS PRECONGL GD IS V SIM TO 30207
: 230209GG04TO CLASTS IN CONGLOMERATE FOUND YESTERDAY
: 230209DY09BOTH BEEN META. THIS PLUTON IS PROB PRECONGL
:mts
```

Figure 6. Retrieval of all comments on conglomerate through searching everywhere in each line for the character string "CONGL".

The subspecs file permits listing of all information concerning a rock unit or specific types of information can be retrieved by simple character-string searches. Conditional searches require short FORTRAN programs, as do calculations on certain specified data and the production of special tables.

Considerable progress has been made in developing editing routines, but a general, verification-edit routine is not presently operational. The SUBMON program was developed with file maintenance in mind. Editing can be done on any sub-file retrieved. The corrected data are then consigned, by one command, to an update file. Before searching the master file, the program goes first to an update file and if it finds the station in question, then it bypasses the master file containing the uncorrected data. When the update file is large enough, it is incorporated into the master file (the program to do this, however, remains to be written).

Use of SARS. The basic concept of SARS is that it allows the field geologist to mimic the procedure he would normally use in compilation. Through SARS he may retrieve the complete data base for any map unit or sub-area as part of that map unit. He can then interrogate the sub-file for all data and comments concerning such topics as lithology, mineralogy, structure, grade of metamorphism, migmatites, dykes, relationships to other units, etc. From this point he can summarize the data, edit the data or map boundaries and then re-compile rapidly. The technique is basically very powerful, fast and flexible. Furthermore, it contains the logic of a cartographic analysis system.

Examples of retrieval follow. Figure 2 illustrates the start of the retrieval of the complete data base for a map unit, namely the TOBA QUARTZ DIORITE – GRANODIORITE, and Figure 3 is an example of the tabular summary of the main features of this sub-area. Comments on dykes selected by retrieving on DY in column 8 are listed in Figure 4. Inter-station comments appear in Figure 5 and illustrate the value of such textural information which would be lost using codes alone. Character-string searches anywhere within free text are also possible (Figure 6).

Conclusions

The following conclusions continue to be valid. (a) The system requires less time on outcrop; (b) less information is lost through illegible handwriting; (c) consistency of recording information is greatly improved because the preprinted sheets serve as checklists; and (d) manipulation and display of geological data can be accomplished in many more ways, both for analysis and publication, than has hitherto been thought practical.

Efficiency is greatly reduced by time consumed on editing of input data, by programming efforts on part of geologists and a deceptive amount of effort required for file and data-base management. Hopefully the pitfalls and progress of this project will increase the efficiency of successor projects of this nature.

Bibliography

Hutchison, W. W., and Roddick, J. A.
1968 Recording geologic field data for machine retrieval and processing: Western Miner, 41, p. 39-43.

Roddick, J. A., and Hutchison, W. W.
1972 A computer-based system for geological field data on the Coast Mountains Project, British Columbia, Canada: 24th Inter. Geol. Congress, Section 16, p. 36-46.

Application of DASCH system to geological field data and documentation files of the Geological Survey, Federal Republic of Germany

R. Vinken
Niedersaechsisches Landesamt
fuer Bodenforschung
3 HANNOVER - BUCHOLZ
P. O. B. 230153
FEDERAL REPUBLIC OF GERMANY

The computer-based system for geological field data currently in use at the Bundesanstalt fuer Bodenforschung (Federal Geological Survey) and the Niedersaechsisches Landesamt fuer Bodenforschung (Regional Geological Survey of Lower Saxony) comprises the three components described below:

Recording System for Geological Field Work Data

The first step towards automatic data processing was to compile a symbol code for a complete and, if necessary, detailed description of geological objects (Look and Vinken, 1971). In fixed sequence but in free format of variable length, the symbols contain the descriptions of the following main groups:

1. Stratigraphy (litho- and biostratigraphy; Precambrian to Quaternary, alpine and non-alpine terms as used in the Federal Republic of Germany)
2. Lithology (sedimentary, magmatic and metamorphic rocks; petrographic parameters)
3. Genesis
4. Colour
5. Additional descriptive characters (e.g. tectonics; structural elements as bedding types, texture; paleontological data, etc.)
6. Sampling

Besides these main groups there appear the usual geographical data, thickness or depth of strata, author, etc.

Each symbol consists of a combination of capital and non-capital letters, figures and special characters. The symbols chosen for the code have been used with slight differences by many of the geoscientists of most of the Geological Surveys in Germany and are known to the map users, as they have been printed in a similar way on the traditional geological maps. The heirarchy and the logic of the symbols facilitate memorizing and permit the presentation of quantitative data.

The code contains, at present, about 4 000 symbols. It is open for ammendments, additions and deletions. The code was published by the Federal Geological Survey together with the Regional Surveys. A second edition of the code of 1971 will appear in 1974. Newly introduced will be symbols from the fields of marine geology, paleontology and engineering geology.

For practical work field forms based on the symbol code system have been introduced. The annual number of objects documented is at present around 10 000, mostly boreholes for mapping purposes with a depth between 2 and 25 m, outcrops, deep boreholes of the oil companies and some hundreds of boreholes with a depth of some 100 m for groundwater supply and engineering geology purposes. The storage of all geological objects of the two Surveys in a standardized data file has been started.

Up to now the data have been stored on punch cards as data carriers. Following installation of our new computer, a SIEMENS 4004 (128 K-bytes), storage on magnetic tape and disc is envisaged. The structure of the data on the data carrier has the same order as within the field recording system described. Two punch-card formats are used: The title card carries information in fixed format on sheet number, name of geologist, outcrop number, coordinates, etc. The description cards carry in free format the stratigraphic descriptions in the same order as on the field forms: stratigraphy up to sampling, and where possible sampling results.

According to our experience the use of the free format system with an alphanumeric code is an absolutely necessary precondition to enable the geologist to record and to document the great variety of geological parameters.

Documentation and Retrieval System DASCH

The documentation and retrieval system with the acronym DASCH (Dokumentations-und AbfragesystemfürSchichtenverzeichnisse, Mundry, 1973) contains 17 subprograms together with about 1200 FORTRAN commands. The necessary core capacity is 128 K-bytes. The program-package is relatively easy to adapt for different computers; at present it is running at a UNIVAC 1108, a CDC Cyber 73 and a SIEMENS 4004.

The documentation part of DASCH supplies a list of the sections or stratigraphic descriptions originally written in free format in fixed columns under the headings: depth interval, stratigraphy, lithology, etc. In addition to the lists, a high-speed printer plots the distribution of all records listed. A program for the decoding of the symbols and the printing in free text will be at hand in 1974. This complete text will enable customers of the Surveys and laymen to read and understand the results more easily.

The second part of DASCH concerns scanning and data retrieval. The scan commands or queries are formed with the same symbols which were used for the description of the geological sections. The structure of the queries is nearly the same as the structure of the field recording system and as the data structure on the data carriers. Therefore it is rather easy for each geologist familiar with his own problems to formulate the queries in order to reach his goal.

DASCH operates sequentially, with a first screening of the bulk of data according to sheet number, coordinates, author, etc. and is therefore useful. Logical operators in the system are AND and OR, the relation operators are GREATER THAN, LESS THAN and FROM-TO. The hierarchical construction of the symbols is used in the scanning for prefixes. For each of the 6 main symbol groups (stratigraphy, lithology etc.) up to 20 queries, each consisting of a maximum of 200 symbols that can be dealt with simultaneously.

As a result of the queries, a retrieval list is printed which contains the retrieved stratigraphic descriptions and an identification character according to the specific questions. At the end a statistical summary of the absolute and percentage number of strata found for each query is printed. A report on the number of strata not considered is given. A first plot shows the distribution of all investigated points considered and a second one indicates by - and + signs those boreholes or outcrops which have no stratum or at least one stratum which satisfies a query. By means of the statistics and of the plots the operating geologist is able to immediately control his commands and the results. For further processing the identification character is coded together with sheet-number, coordinates, etc., in such a way that the use of pure numerical programs is possible.

Further Processing of General Geological Data Delivered by DASCH

Data retrieved from DASCH can be processed by a number of available routines (Mundry, unpublished) as programs for one dimensional statistics, regression analysis, factor analysis, cluster analysis, trend surface analysis, etc., including subprograms for plotting (ZUSE Z 64 Graphomat).

Originally we considered the compilation of a program-package for all aspects connected with the computerization of geological mapping as compared with our field recording system. For further processing, however, the requirements of the users are too manifold; therefore a package proved impractical. The preparation of a set of specialized programs and the access to existing programs was unavoidable. A second aim was to construct geological maps by computer directly from the field data (outcrop and borehole descriptions, geomorphological investigations) in contrast with cartographic reproduction for printing after digitizing manuscripts prepared in a traditional way.

A contouring package is available (CDC Cyber 73, Z 64 plotter) especially for graphical manipulation. For example, for about 30 sheets at a scale of 1 : 25 000 covering the German Northsea coast, the isolines of the base of the marine Holocene sediments are constructed and plotted from all available drillhole descriptions (Barckhausen, in press). Isopach maps are for the total Holocene. Sand layers of the Holocene for the same area are in preparation and the construction of facies maps is in the planning stage.

Other possibilities existing at the moment for the graphical representation of DASCH-retrieved data (Barckhausen, 1973) are the construction and plotting of geological sections directly from field descriptions, taking into account up to four different components (e.g. sand, clay, peat, lime content, or: sandstone, siltstone, limestone, clay content). There is, furthermore, the possibility to represent the distribution of different strata, which in part overlie each other, in so-called symbol-point maps. A program under development will lead to the representation of geological boundaries in these maps. Another development resulting from the existing programs (sections and symbol points) is the construction of geological maps called "profile-type maps"(Hageman, 1962) which may represent up to 10 strata overlying each other and which for laymen are more easily readable, than the conventional geological maps, which have become more and more complicated.

Bibliography

Barckhausen, J.
1973 Automatisch gewonnene Informationen aus geologischen Schichtenverzeichnissen and ihre Weiterverarbeitung: Geo. Jb. A7, S. 35–51, 5 Abb., 3 Tab.

Hageman, B. P.
1962 Anweisung zur Herstellund der Geologischen Karten in Marschengebieten: Niederl. Geo. Dienst (dt. Übersetzug).

Look, E. R., and Vinken, R.
1971 Elektronische Datenverarbeitung bei der Aufnahme and Herstellung von geologischen Karten. Mit einem ausführlichen Anhang: Symbole aus der Geologie für die elektronische Datenverarbeitung, 181 S., 5 Abb., Bundesanstalt für Bodenforschung und Geol. Landesamter der Bundesrepublik Deutschland.

Mundry, E.
1973 Ein Documentations und Abfrageprogramm für Schichtenverzeichnisse (DASCH): Geol. Jb. A 7, S. 25-33, 1 Tab.

Aarhus University System II

S. W. Platou
Laboratoriet for Geofysik
Geologisk Institut
Aarhus Universitet
Finlandsgade 6 - 8200 Aarhus N
DENMARK

Introduction

SYSTEM II is an experimental data-processing system developed to handle geological field and laboratory data. The system is not project-oriented, as a group of standard parameters specified for each project make it possible for the system to handle the data. The system was originally developed to handle data from mapping of Precambrian rocks in West Greenland and during this project the basic recording principles were developed (Platou, 1971). Later the system was used in projects of the following types: Geochemical prospecting for flourite, prospecting for nickel ores, petrological investigation of a gabbro massif, handling of petrophysical properties of granitic rocks, and in mapsheet mapping by the Geological Survey of Sweden, where the system partly has replaced the GEOMAP system (Berner et. al., 1971).

Programs in System II

The main system consists of 6 programs, as shown in Figure 1. Two main programs handle the initial treatment of the data and the administration of the project parameters. Four programs are used to sort, list and depict the data. The programs are written in ALGOL-60 and run on a Control Data 6400 computer with 32 K (decimal) core store available to the user. The programs in the system require 18-28 K core store, of which 8 K is used for the plotter system. ALGOL was chosen as the programming language because of better output facilities and easier handling of logical expressions and DO statements, than in, for example, FORTRAN IV. But to reduce computing time, some of the routines have been written in FORTRAN, as combined FORTRAN/ALGOL programs can be run on the CDC 6400 computer. The computing time required for running the system is mainly dependent on the amount of output wanted, for example, sorting out and printing of 100 datasets from a 4 000 dataset file can be done in 4-5 seconds CP-time, of which 1 second is used to load the program into the computer, and 1-2 seconds are used for the sorting. The small computing time means that the system, to a large extent, can be operated interactively from a terminal and this is done by means of a Tektronix 4010-1 graphical display terminal.

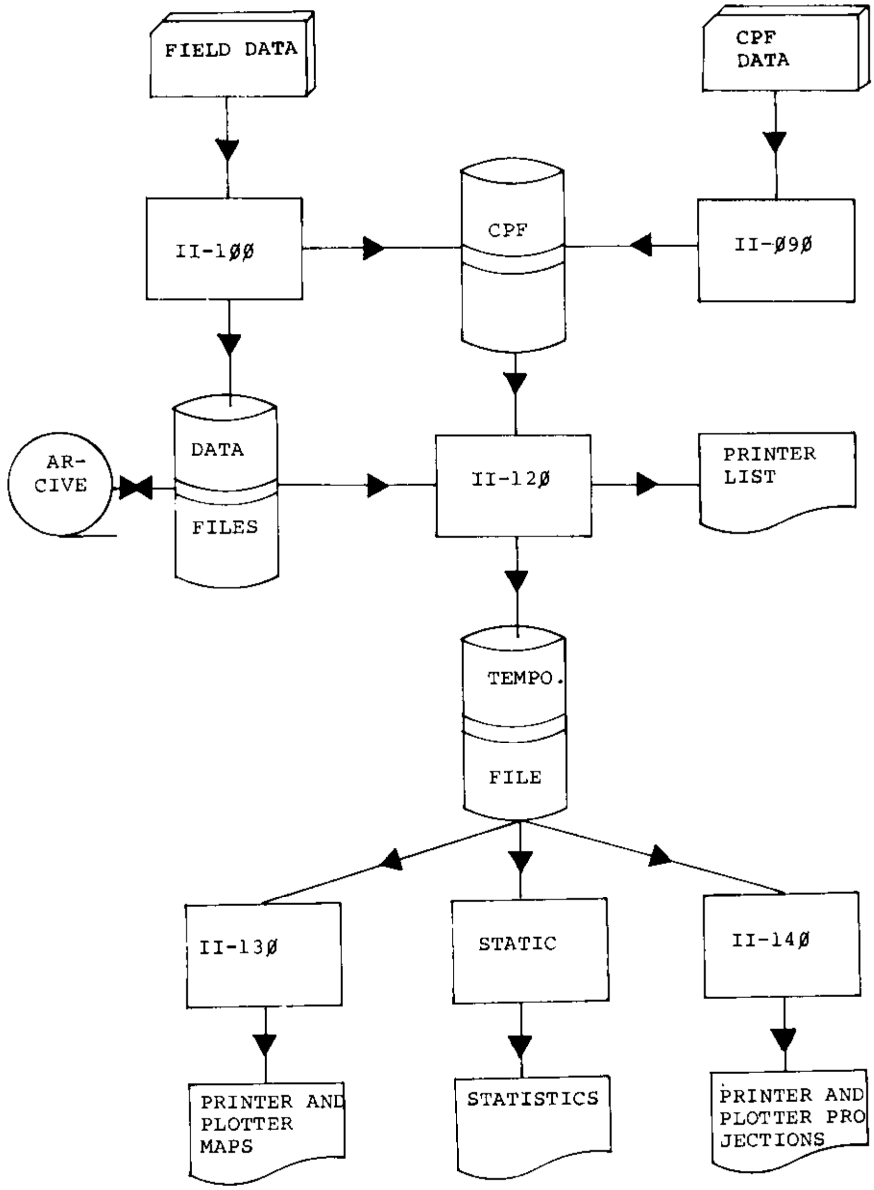


Figure 1. Treatment of field data in System II. The CPF file is created once for each project, and updated when necessary by means of program II-090. The field data is by means of program II-100 checked for errors and transferred to a field data file. The field data files are during active use stored as disc files and archived on magnetic tapes. Program II-120 does the major part of the sorting in the system and the program produces lists with the wanted dataset values. Data to be plotted on maps, treated statistically or plotted in projections is sorted out by means of II-120 and stored on temporary files. The map plots are done by means of program II-130, and the maps can be delivered on printer and Calcomp plotter of graphical display. The statistical calculations are done by means of a standard FORTRAN IV program library. The orientation measurements are partly plotted on maps by means of II-130 and can partly be depicted in projections by means of program II-140. This program can produce plot-of-measurement stereographic projections, contoured stereographic projections and can calculate great and small circles and plot these circles.

User Facilities

The main purpose of developing the system was to create a system which could accept data of any type, and allow the user to create recording sheets, which from the user's point of view are as effective as possible. In practice this means that the user can record what he wants and create recording sheets with very little empty space. This made it necessary to develop completely new recording principles which partly fulfill the geologist's requirements and partly make computer treatment of the data effective. This is only possible by a general use of coding in the datasets and by using standard recording structures, as described by Platou (1971). The extensive use of codes for texts have been found to be of major importance in making possible the fast computer treatment of the data. None of the 30-35 geologists who have used the system, have had any trouble with the codes for three reasons: 1. It takes less time to write a code value than to write a full text; 2. The number of recording errors and punching errors become very small; and 3. The number of codes regularly used, during field work, is small and therefore learned during a couple of days field work.

Laboratory and field data are recorded using the same principles and are handled by the same programs. A number of auxiliary programs are connected with the system to handle petrological calculations and depicting, for example, triangular diagrams. The field and laboratory data can easily be treated together; for example, laboratory and field data may be plotted in the same run.

It was found necessary to be able to use coordinates from any coordinate system, mainly to treat data from local investigations together with more general information from an area. But to make the use of coordinates easier, all coordinates in the data files are in the same system.

The output from the system is: 1. Lists in which the user defines which values it should contain; 2. Printer and plotter maps on which two independent variables, one as symbols and the other as values or text, can be depicted; and 3. Various types of projections with orientation measurements. The problem for the user is to choose an output format, because the system contains about 35 completely different ways of depicting data.

Data Management

The parameters needed by the programs in the system to handle data from a specific project, is stored in a separate file named the Common Project File (CPF). The CPF contains general project parameters, parameters for each of the dataset values in the project, the text corresponding to the codes used and the parameters for each of the data files in the project.

The CPF is fundamental to the ability of the system to do fast processing of the various data groups and files in a project. The content of the CPF allows the user to stipulate coordinates, scales, the values he wants and how he wants to depict the data.

The files treated by the system until now (up to 15 000 datasets) have not required reorganization of the data in the files. A reorganization seems not to be necessary as long as only the geologist himself, or the team working in a project, is using the data. But if the data are to be a part of a larger data base and extraction and merging of data seem necessary, mainly to purge data which are not of interest and to combine data about the same items from various files, it would, for example, be possibe to combine field and laboratory data for various rock types and from several projects. Facilities for doing this are included in the system, because it is easy to create new recording formats similar to field sheets and to extract data from various files and merge them into such new formats.

File Structure

Both data and file structures are rather complex, but this is not noticed by the user because the programs by means of the CPT, do all the administration. The user need only know the names of the files he wants to use.

The content of a data file is a combination of three hierarchial systems which are all meant to make the sorting and datafile handling effective. The three systems are: 1. The recording system; 2. The data file system; and 3. A sorting system. It is nearly impossible to briefly describe the various hierarchies, mainly because each project has its own combination and because one level can be of major importance in one project and not even used in another project.

Future Development and Conclusions

A problem with System II is its many possible uses: one can record whatever is wanted and the data can be handled and depicted in a large number of different ways. But it is not unimportant how the recording is done in relation to what the user wants. A typical example is that the user wants to carry out intensive statistical treatment on his data, but he had not classified the data in such a way, during the recording, to make it possible for the computer to separate the various data groups in a simple way. Another serious problem is to make the records useable for those other then the geologist himself. The major problem in the future seems, therefore, not to be a further sophistication of the computer programs, but an investigation of the relation between recording techniques and what the data should be used for. A geological map sheet is a good example of these problems. A geologists collect data, treat them by computer and produce a map. Another user takes the map and extracts the data he needs for his purpose, but much is lost in this process, especially the possibility for judging the reliability of the data depicted on the map. It would be much better if the outside user could ask a data base for the information he needs. But it is unfortunately rather unknown what such a data base should contain, mainly because of lack of information from the outside users about what they would like to have in such a data base.

Another problem is the education of the geologist in the use of computers because it does not seem to be generally realized that the effective use of a computer requires just as much education as is the case in the use of a microscope. Most geologists use computers to do the same work they do by hand, and only very few utilize the new possibilities offered.

References

Berner, H., Ekström, T., Lilljequist, R., Stephansson, O., and Wikström, A.
1971 Data storage and processing in geological mapping. I: Field data sheet II: Data file. Geol. Före. Stockholm Förh., 93, p. 85-101 and p. 693-705.

Platou, S. W.
1971 An electronic data-processing system for geological field and laboratory data. The E.D.P. system Agto., Geological Survey of Greenland, Report no. 39, 42 p.

PART 2: GENERALIZED GEOLOGICAL DATA MANAGEMENT SYSTEMS

Introduction

G. D. Williams
Department of Geology
University of Alberta
Edmonton, Alberta T6G 2E1
CANADA

Geological field data comprise an exceptionally broad spectrum of observations, analytical results, inferences and conclusions relating to points, linear features and areas of the crust, collected under diverse conditions, often by many individuals over a considerable time span. Because of the enormity of the task before him, the field geologist usually functions as an analog device, interpreting his observations as he makes them, analyzing the consequences, and often recording only the conclusions, usually in map, or other graphic form. Each individual performing this task reflects in the conclusions he draws, his training, experience and personal biases as well as the influence of local conditions under which he is operating. Many field data, therefore, have been considered to be so interpretive, so individualistic and so lacking in reproducibility that at best, special procedures are necessary in order to deal with them successfully in computer files, or at worst, computer processing should not be used because it destroys the unique relationship between the geologist, his field area and his conclusions by requiring unnatural rigidity in his operations.

Geological field data, however, form the foundation upon which is built virtually all of our knowledge of the earth. Because of this importance, ways must be found not only to collect it more efficiently and objectively, but to use it more effectively in the service of man and science: in the study and utilization of man's environment and in providing for his needs in the form of energy, minerals and renewable resources, and in solving problems related to the origin and history of the earth and our universe. These are areas of concern which have been identified specifically by UNESCO and IUGS and upon which attention has been focussed through the International Geological Correlation Programme.

Efficient, effective and timely use of geological field data now is more essential than has ever been the case in the past, and can only be accomplished by application of modern techniques of data management and analysis. The papers which follow describe generalized computer systems which have been designed and used for managing geological data of diverse types, including field data. Whereas geological field data are undoubtedly unique in many aspects, the problems of managing the data, and to a large extent analyzing them are not, and can be most efficiently solved by application of generalized, rather than ad hoc techniques.

Le Système SIGMI de l'Ecole Nationale Supérieure des Mines de Paris

M. Kremer, M. Lenci, et M.-T. Lesage
Ecole Nationale Supérieure des Mines de Paris
35, Rue Saint Honoré
77 Fontainebleau
FRANCE

Introduction

Le système de Gestion de Base de Données "SIGMI" a été élaboré dans le cadre d'un programme de recherche visant à améliorer le traitement des données géologiques à l'aide des ordinateurs. Ce système répond à la nécessité de disposer d'un outil généralisé capable d'effectuer les opérations élémentaires sur les fichiers sans programmation préalable, permettant ainsi de focaliser les efforts sur les véritables exploitations de ces données et non sur leur gestion.

Les principaux objectifs qui ont guidé la réalisation de ce système ont été:

1. Il doit être possible d'utiliser les mots du langage naturel à tous les niveaux, les codifications devenant ainsi facultatives.
2. Le système doit être orienté vers l'utilisateur, c'est à dire qu'il est possible à des géologues non-informaticiens d'avoir directement accès aux données, indépendamment des utilisations qu'ils en feront.
3. Le contenu et la structure des enregistrements doivent pouvoir évoluer en fonction des conditions de la collecte ou des besoins des utilisateurs.
4. Le système doit permettre le traitement des données non numériques à l'aide de codes sémantiques.
5. Efin le système doit pouvoir traiter aussi bien des fichiers documentaires que des fichiers informatifs, ce qui conduit à abondonner la distinction traditionnelle entre système d'information et système de documentation.

Principes Généraux

Le format d'entrée. Le système SIGMI utilise un format libre tel que celui utilisé dans certains systèmes de documentation. Le principe est simple, il suffit, pour chaque donnée que l'on veut stocker, d'expliciter le nom de l'élément (ou item) auquel cette donnée appartient.

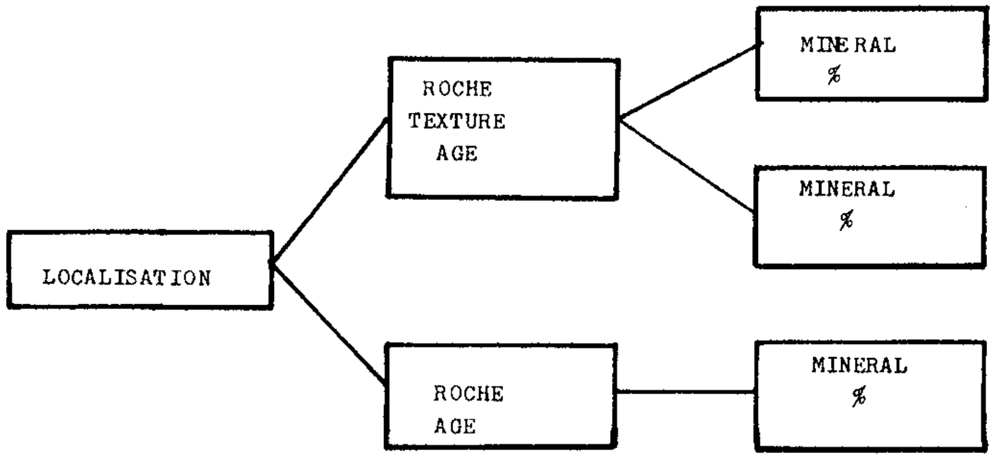


Figure 1.

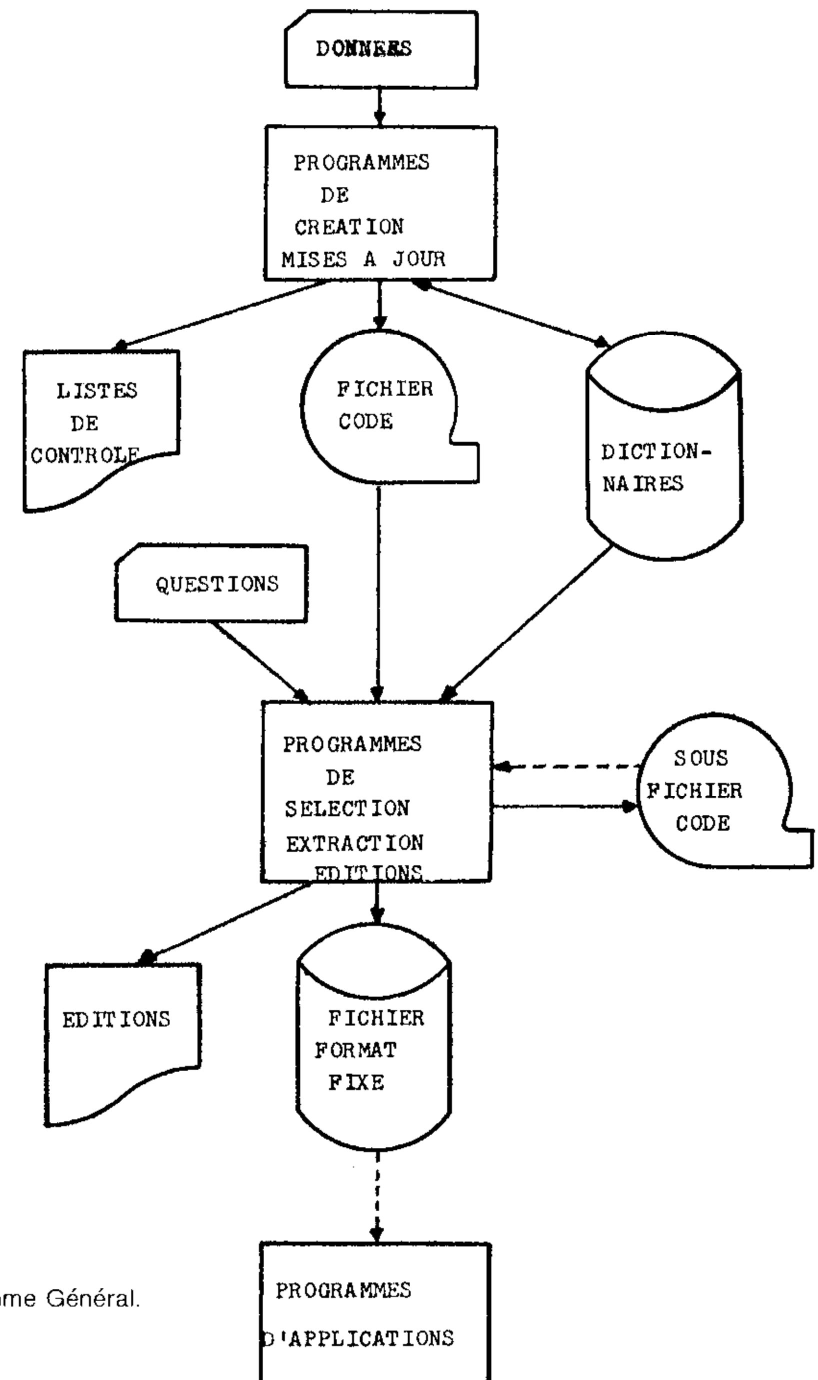


Figure 2. Organigramme Général.

Exemple:

ROCHE = GRANITE/MINERAL = QUARTZ/MICA/$

Trois caractères séparateurs sont utilisés:

- pour séparer le nom de l'élément de ses valeurs: "=".
- pour séparer les couples élément-valeur, ou les valeurs d'un même élément: "/".
- pour séparer les enregistrements: "$".

Cette méthode permet un stockage réellement libre des données permettant entre autres:

- l'utilisation du langage naturel
- le stockage des données dans un ordre quelconque
- un nombre d'occurrences des données pratiquement illimité.

Les types de données. En fonction de leurs analogies conceptuelles, les données sont réparties en un certain nombre de "types". A chaque type correspond un mode de stockage et des traitements particuliers dont le but est de rendre plus efficace et rigoureux l'exploitation des données. Les six types actuellement définis sont:

- Type numérique standard: un nombre par valeur.
  Exemple: SIO2 = 52,4/
- Type numérique standard: un nombre par valeur.
  Exemple: PUISSANCE = 10 A 15/
- Type coordonnées géographiques: deux valeurs (point) ou quatre valeurs (surface) séparées par une virgule.
  Exemple: COORDONNEES = N8D 58M, W4D 52M15S/
- Type avec codage interne standard.
- Type commentaire: sans codage, la recherche s'effectuant seulement sous forme de chaine de caractères.
- Type sémantique: présence pour ces éléments de dictionnaires de codification sémantique.

La structuration des enregistrements. L'inconvénient de ce format d'entrée résidait dans la difficulté de structurer les enregistrements, ce qui était absolument indispensable pour un système de Base de Données. La méthode la plus naturelle, qui a été employée, consiste à relier les couples élément-valeur, précédemment définis, par des jeux de parenthèses imbriquées.

Exemple: LOCALISATION = X(ROCHE = GRANITE/TEXTURE = GRENUE/AGE =
HERCYNIEN(MINERAL = QUARTZ/% = 30)(MINERAL = MICA/% = 5))
(ROCHE = GRES/AGE = EOCENE (MINERAL = QUARTZ/% = 95))$

Cette écriture est en fait une représentation linéaire de structure arborescente et l'enregistrement ainsi constitué peut être représenté par un arbre où chaque passage à niveau supérieur correspond à l'ouverture d'une parenthèse (Fig. 1). Ce procédé de stockage permet de n'avoir à déclarer ni le contenu, ni la structure des enregistrements avant leur introduction dans le fichier, chaque fiche constituant un tout se décrivant lui-même.

Le Langage d'interrogation

Le langage d'interrogation comporte six paragraphes permettant de réaliser trois fonctions:

1. Sélectionner les enregistrements d'un fichier répondant à un certain nombre de questions: paragraphes CRITERES, STRUCTURES et LOGIQUE.
2. Transférer les enregistrements sélectionnés sur un sous-fichier de même structure que le fichier d'origine: paragraphe SOUS-FICHIER.
3. Extraire certaines données des enregistrements et constituer un sous-fichier de format fixe pouvant servir ultérieurement à des programmes d'applications: paragraphes VARIABLES A EXTRAIRE et ASSOCIATIONS DES VARIABLES. Ce langage d'interrogation, dont nous donnons un exemple en Annexe I, présente quelques particularités:

(a) A chaque type de données correspond un jeu particulier d'opérateurs dans le paragraphe CRITERES.

Exemples: - SIO2<45: signifie, SIO2 inférieur à 45.

- PAYS <AFRIQUE: signifie, PAYS inclus dans AFRIQUE.

(b) Les paragraphe STRUCTURES et ASSOCIATIONS DES VARIABLES sont indispensables du fait qu'un même élément peut être répété plusieurs fois et apparaître à des positions différentes dans l'arbre. Les opérateurs utilisés fixent la position relative des critères dans l'enregistrement à la manière d'un arbre généalogique (jumeau, frère, père, fils, etc.).

Exemple: C1.JU.C2 (soit C1 jumeau de C2): signifie que les deux critères doivent se trouver réalisés à un même noeud dans l'arbre.

Principales Caractéristiques: (Fig. 2)

Mode d'utilisation: batch processing
Nombre de programmes: 10
Langage: COBOL et quelques sous-programmes en ASSEMBLEUR 360
Nombre d'instructions: environ 10 000
Ordinateurs: IBM 360 et 370, CII IRIS 80
Ressource mémoire: 120 K octets
Périphériques: 1 disque et 2 bandes magnétiques

Annexe 1

Exemple d'Interrogation du Fichier "Roches Ignées"

Critéres

C1: PAYS<AFRIQUE/
C2: MINERAL<FELDSPATHOIDE/
C3: MINERAL<SILICATE + FE + MG/
C4: MINERAL<FELDSPATH/
C5: MINERAL<HORNBLENDE/

C6: MINERAL<BIOTITE/
C7: MINERAL<OXYDE + FE/
C8: SIO2>39/
C9: SIO2<45/
C10: NA2O> = 5/
C11: CAO> = 5/

Structures

S1: C8.ID.C9/
S2: S1.JU.C10.JU.C11/

Logique

C1 ET C2 ET C3 ET NON C4 ET (C5 OU C6 OU C7) ET S2/

Sous-Fichier

OUI/

Variables à Extraire

V1: GEOL (I6)/
V2: SIO2 (F7.2)/
V3: AL2O3 (F7.2)/
V4: NA2O (F7.2)/
V5: CAO (F7.2)/
V6: K2O (F7.2)/
V7: NI (E9.2) F/
V8: CO (E9.2) F/

Associations des Variables

A1: V2.JU.V3.JU.V4.JU.V5.JU.V6.JU.V7.JU.V8/
A2: A1.FI.V1/

Bibliographie

Kremer, M.
1972 Essai de structuration de données dans un fichier d'archive géologique à format libre: Rapport interne E.M.P., CIG/R72/1.

Kremer, M., Lenci, M. T.
1973 SIGMI: un nouveau système d'interrogation de fichiers orienté vers l'utilisateur: Rapport interne E.M.P., CIGD/R73/5.

Kremer, M. et Lesage, M. T.
1972 Système d'information géologique et minière: exploitation d'un fichier "Cartes et Plans": in Traité d'Informatique Géologique à la dir. de P. Laffitte, Masson, Paris.

Laffitte, P.
1969 La codification sémantique en informatique géologique: Annales des Mines, Déc. 69, p. 75-83.

Lenci, M.
1972 Notions de base de la codification sémantique: Rapport interne E.M.P., COG/R72/4.

Lesage, M. T.
1970 Programme SIGMI: Fichiers Cartes et Plans: Rapport interne E.M.P., CIG/R70/5.

SIGMI system of the National School of Mines of Paris

M. Kremer, M. Lenci et M.-T. Lesage
Ecole Nationale Supérieure des Mines de Paris
35, Rue Saint Honoré
77 Fontainebleau
FRANCE

Introduction

The SIGMI Data Base Management System has been developed within the framework of a research programme whose aim was to improve the processing of geological data with the help of computers. The system answers the requirements that are necessary for a generalized tool to be able to perform the basic operations on files without any preliminary programming. This allows efforts to be focused upon the actual exploitation of the geological data and not upon data processing. The principal aims which have guided the development of this system are:

1. It must be possible to use natural language at all levels.
2. It has to be a user-oriented system so that the geologists who are not particularly familiar with information processing may directly access the data, independently of use.
3. The content and structure of the records must be able to be continually developed according to the conditions of data gathering or the needs of the users.
4. The system must allow processing of non-numerical data with the help of semantic codes.
5. And finally, the system must be able to process documentation files as well as data files. With this idea, the traditional distinction between an information system and a documentation system is abandoned.

General Principles

Input Format. The SIGMI system uses free format such as that used in some documentation systems. The method is simple: for each piece of data to be stored, the name of the element for items to which the data value belongs is specified. For example:

ROCK=GRANITE/MINERAL=QUARTZ/MICA/$

Three characters are used as separators:

- in order to separate the name of the element from its values: "=".
- in order to separate the pairs "element-value" or the values of a

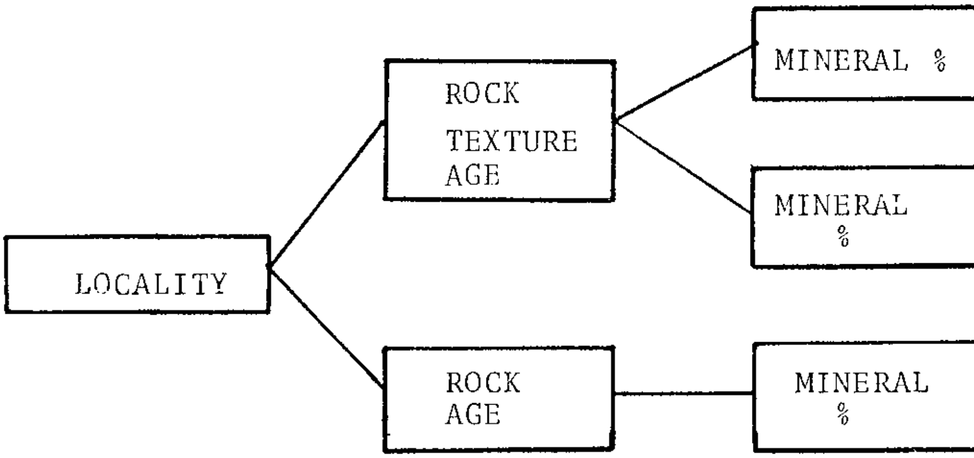


Figure 1.

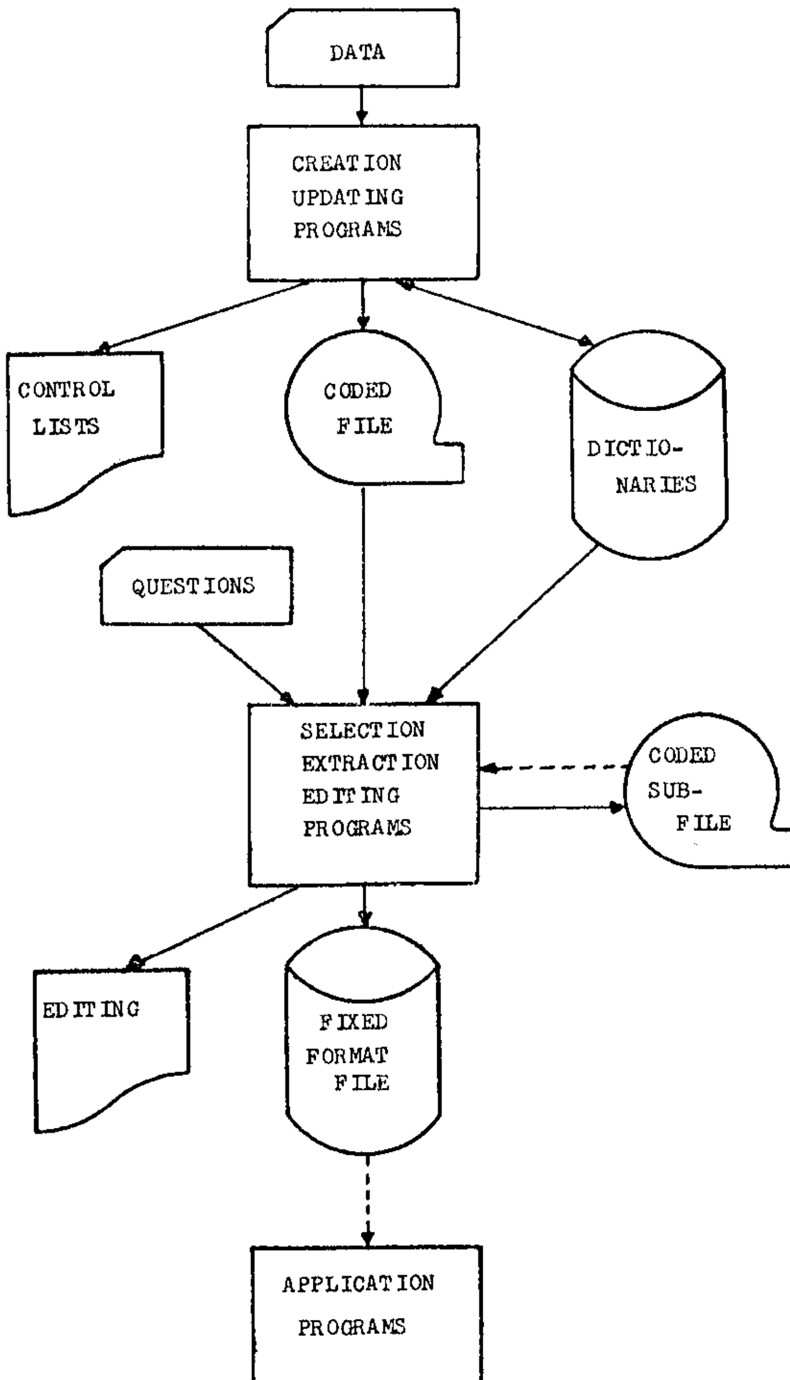


Figure 2.

same element: "/".
- in order to separate the records: "$".

This method allows an absolutely independant storage of the data permitting:

- use of natural language words
- data storage in any order
- a practically unlimited number of data occurrences

Data types. According to their conceptual analogies, the data are divided into a certain number of "types". To each type there corresponds a special way of storage and processes so that the data operations are more efficient and more rigorous.

The six types actually defined are:

- standard numerical type: a number per value
  For example: SIO2 = 52,4/
- numerical type with two limits: two numbers per value
  For example: THICKNESS = 10 TO 15/
- geographic coordinates: two values (point) or four values (surface) separated by a comma. For example: COORDINATES = N8D 58M, W4D 52M15S/
- type with an internal standard coding
- commentary type: no coding so the search is only made on character strings
- semantic type: for these elements, existence of dictionaries dealing with semantic coding.

Record structure. The draw-back of this input format could be found in the difficulty of structuring the records which is absolutely necessary for a data base system. The following method has been used: it consists of linking together the value-element couples which have been formerly defined by imbricated brackets. For example:

LOCALITY=X(ROCK=GRANITE/TEXTURE=GRANULAR/
AGE=HERCYNIAN(MINERAL=QUARTZ/%=30)(MINERAL=MICA/%=5))
(ROCK=SANDSTONE/AGE=EOCENE(MINERAL=QUARTZ/%=95))$

This writing is, in fact, a linear representation of an arborescent structure and the record thus realized may be represented by a tree in which each entry to a higher level corresponds to the opening of a bracket (Figure 1). This method of storage does not require the declaration of the content nor the record structure before their introduction into the file, each data sheet being an entity which describes itself.

Interrogation Language

The interrogation language is composed of six paragraphs allowing three functions:

1. The selection of file records answering a certain number of questions: CRITERIA, STRUCTURE and LOGIC paragraphs.

2. The transfer of the selected records into a sub-file of the same structure as the original file: SUB-FILE Paragraph.
3. The extraction of certain data from records and the constitution of a fixed format sub-file which may be used at a later date for application programs: VARIABLES TO BE EXTRACTED and ASSOCIATIONS OF VARIABLES. This interrogation language, an example of which is given in Appendix 1, shows a few particular points:

   (a) To each type of data corresponds a particular set of operators in the CRITERIA paragraph. For example:

   - SIO2<45: means, SIO2 less than 45.
   - COUNTRY AFRICA:< means, COUNTRY included in AFRICA.

   (b) The paragraphs STRUCTURE and ASSOCIATIONS OF VARIABLES are necessary owing to the fact that the same element may be repeated several times and appear on different positions in the tree. The operators used fix the relative position of criteria in the record as in a genealogical tree (twin, brother, father, etc.). For example:

   C1.JU.C2 (C1 twin of C2): this means that both criteria have to occur at the same node in the tree.

Main Characteristics: (Figure 2)

Processing method: batch processing
Number of programs: 10
Language: COBOL and a few sub-programs in Assembler 360
Number of instructions: about 10 000
Computers: IBM 360 and 370, CII IRIS 80
Core storage space: 120 K bytes
Peripherals: 1 disc and 2 magnetic tapes

Appendix 1

Example of Interrogation of "Igneous Rocks" File

Criteria

C1: COUNTRY<AFRICA/
C2: MINERAL<FELDSPATHOID/
C3: MINERAL<SILICATE + FE + MG/
C4: MINERAL<FELDSPATH/
C5: MINERAL<HORNBLENDE/
C6: MINERAL<BIOTITE/
C7: MINERAL<OXIDE + FE/
C8: SIO2>39/
C9: SIO2<45/
C10: NA2O> = 5/
C11: CAO> = 5/

Structure

S1: C8.ID.C9/

S2: S1.JU.C10,JY.C11/

Logic

C1 AND C2 AND C3 AND NOT C4 AND (C5 OR C6 OR C7) AND S2/

Sub-File

YES/

Variables to be Extracted

V1: GEOL (16)/
V2: SIO2 (F7.2)/
V3: AL2O3 (F7.2)/
V4: NA2O (F7.2)/
V5: CAO (F7.2)/
V6: K2O (F7.2)/
V7: NI (E9.2) F/
V8: CO (E9.2) F/

Associations of Variables

A1: V2,JU.V3.JU.V4.JU.V5.JU.V6.JU.V7.JU.V8/
A2: A1.FI.V1/

Bibliography

Kremer, M.
1972 Essai de structuration de données dans un fichier d'archive géologique à format libre: Rapport interne E.M.P., COB/R72/1.

Kremer, M., Lenci, M. et Lesage, M. T.
1973 SIGMI: un nouveau système d'interrogation de fichiers orienté vers l'utilisateur: Rapport interne E.M.P., CIG/R73/5.

Kremer, M. et Lesage, M. T.
1972 Système d'information géologique et minière: exploitation d'un fichier "Cartes et Plans": in Traité d'Informatique Géologique sous la direction de P. Laffitte, Masson, Paris.

Laffitte, P.
1969 La codification sémantique en informatique géologique: Annales des Mines, décembre 1969, p. 75-83.

Lenci, M.
1972 Notions de base de la codification sémantique: Rapport interne E.M.P., CIG/R72/4.

Lesage, M.-T.
1970 Programme SIGMI: Fichiers Cartes et Plans: Rapport interne E.M.P., CIG/R70/5.

G-EXEC: a generalized FORTRAN system for data handling

Keith G. Jeffrey
Institute of Geological Sciences
Exhibition Road
London SW7 2DE
UNITED KINGDOM

Elizabeth M. Gill
Atlas Computer Laboratory
Science Research Council
Chilton, Didcot
Berkshire
UNITED KINGDOM

During the spring and summer of 1972, staff of the IGS Computer Unit analyzed IGS data handling needs. These were formulated into a set of objectives, and against these goals some currently available data handling systems were evaluated. This study resulted in a decision to develop our own data handling system to meet these same objectives.

The objectives were that the system should handle almost any kind of data (generalization) and that the system should be 'portable' (run on any medium-sized computer). We also required integration through data management, data analysis and data display and the system had to be amenable to modification and extension by programmers. Finally, the system had to handle 'project' data and 'archival' data with equal ease.

To meet these objectives G-EXEC was designed to be highly modular. This allows ready building, amendment and updating of program segments. It also saves on storage of the object code in the system program library. The software is written in FORTRAN IV so that it may be handled by most computers. The system is implemented from this 'pool' of modules, and the main (central) system uses the same modules as the small (satellite) subsystems. Furthermore, this technique allows experimentation in different data management or analysis methods without rewriting the software.

The files in G-EXEC are normal computer files, as may be handled by a normal FORTRAN program. There are no constraints on the input in terms of field positions and there is no need for special information to allow the system to structure the data. Single files may be added to the system and may or may not enter the data base. All file linkages are externally defined (in the present implementation) and so great flexibility is achieved. The constraints on the files are that

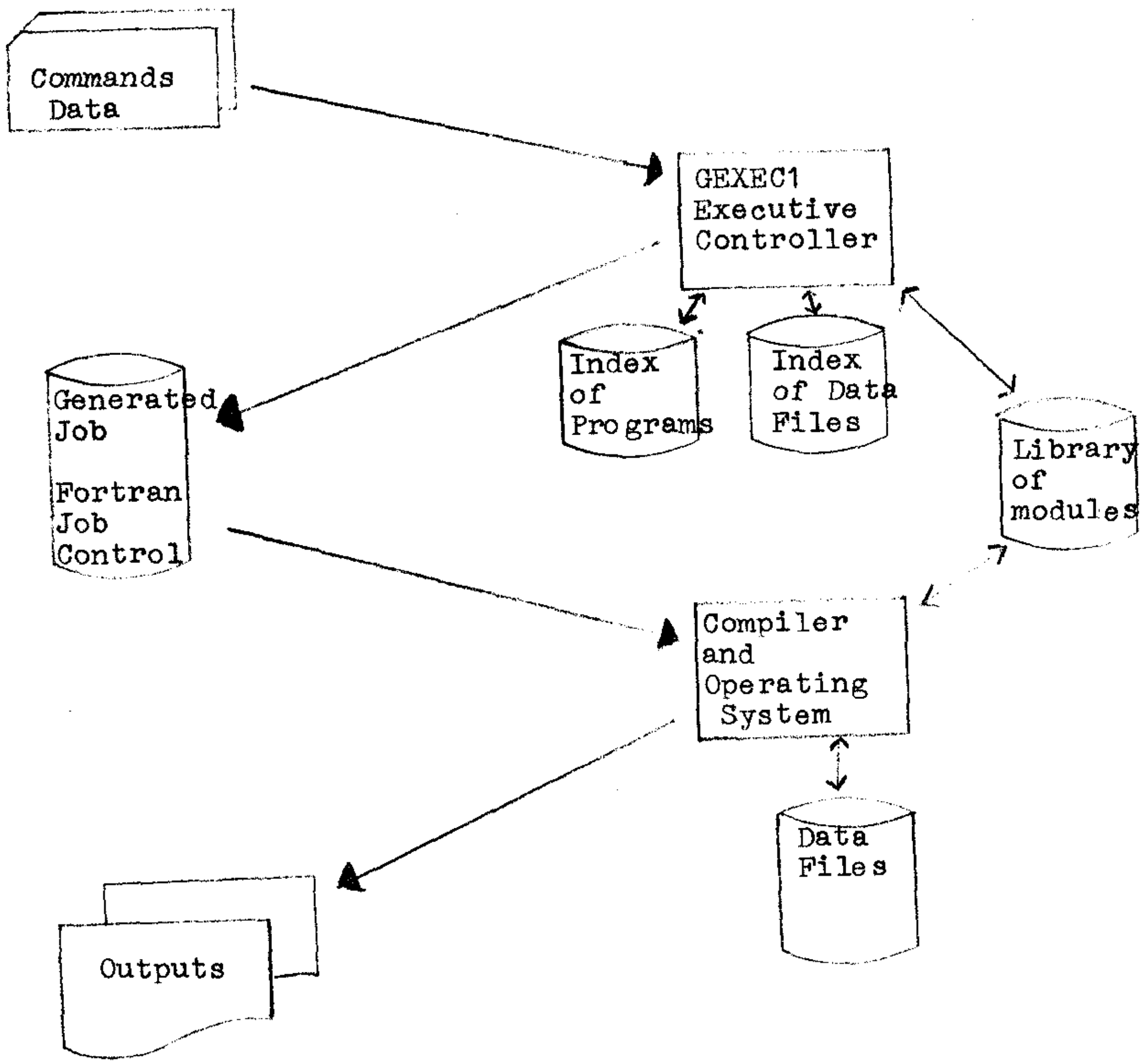


Figure 1. G-EXEC: a generalized FORTRAN system for data handling. The two-stage execution of a G-EXEC Job under Phase 2 of the project. The Executive controller, GEXEC1, compiles commands into a job consisting of FORTRAN and Job Control Language. This job is submitted to the compiler and operating system as a normal batch job. It acts on data files, and on loading incorporates software from the object module library.

all records must contain the same fields, and that each field must be of the same length and type throughout all records in the file. This standard file allows integration.

The other aspect of file design is concerned with generalization. A data description of standard form is added to each file. Programs processing the file access the data in the file through the data description, and so are data independent. The data description contains information on the name, length, type and input format of a field vector, together with its maximum and minimum value and a code representing absent data as opposed to zero value in that field. The data description also allows a dictionary segment to be specified for each field vector to allow coding or decoding to or from the stored form.

The design of the software (highly modular) and of the files (vector storage) both allow for efficient processing, and look to fourth generation techniques.

The implementation of the system on the IBM 370/195 at Rutherford High Energy Laboratory, and accessed via the Atlas Computer Laboratory, has proceeded in planned phases. In Phase 1 the system was largely in the current state of the satellite subsystems, requiring a small amount of FORTRAN programming (typically 5 statements) and some Job Control Language to specify the files required to the operating system. In Phase 2 a modular Executive Controller has been written. This controller is slightly machine dependent, and some parts would need differently coded modules on other computers. However, this could be made general at the expense of run-time efficiency. Under this controller, the user only inputs simple commands, and the software 'compiles' the commands into a job consisting of FORTRAN and Job Control statements. This job then actually processes the task (Figure 1).

Throughout the G-EXEC System there are three faces. The user sees a system with capability to handle his data, and with which he interacts by simple commands. The programmer sees a collection of modules, and can build new applications software or subsystems which can be accessed by users. The system manager sees an audit trail of what the system has done, and he can keep records of users, the processing programs they used and the files they accessed. Furthermore, he has a report of the resources used by each task, ready for an accounting run at periodic intervals.

In conclusion, a few figures are provided to show the system in perspective. The software consists of 28 000 lines of source code, split among some 180 modules. There are, at present, over 100 temporary work-files stored on the system, occupying 50 megabytes, and at any one time there are up to 200 files using the system but not being stored, representing another 20 megabytes. Data files currently being processed by the system on an experimental basis concern petrology, paleontology, regional geochemistry, mineral assessment, structural geology, mineral trade and production statistics, hydrogeology, lithology and stratigraphy (both surface and boreholes), geotechnical properties and sedimentary structures, as well as personnel and accounting files. The system uses 1 IBM 100 megabyte disc pack as the 'active area' for both programs and data. The current workload averages 100 seconds per day on the IBM 370/195.

We should like to thank our colleagues for their constructive criticism and hope that the software we are now developing will eliminate some of the deficiencies in the system.

U.W.O. SAFRAS system

P. G. Sutterlin
Department of Geology
University of Western Ontario
London, Ontario N6A 3K7
CANADA

Introduction

The SAFRAS system is the direct result of a cooperative project initiated in the spring of 1968. At that time, the Ontario Department of Mines (now the Division of Mines, Ministry of Natural Resources) and the Department of Geology of the University of Western Ontario, on the basis of data which had been collected on silver deposits in Ontario, agreed to build a computer-processable file of these data. The file design and programming functions were to be carried out at the Univeristy under a Special Grant of the Geological Survey of Canada. The Special Grants themselves were a result of the recommendations of an ad hoc Committee of the National Advisory Committee on Research in the Geological Sciences (Brisbin and Ediger, 1967), the intent of which was to promote research in support of the National System concept enunciated by the ad hoc Committee.

It soon became apparent that, with a slight shift of emphasis, the project could result in the development of a number of computer programs which could be applied to sets of data other than just those pertaining to the Ontario silver deposits. It was then agreed that the project would concentrate on the development of a system to store, edit, selectively retrieve and otherwise manipulate a range of geological data by computer. It was also stipulated that the system be as user-oriented as possible and that it be relatively easily transferable from one type of computer to another. It was with these considerations in mind that the SAFRAS system was designed and developed.

SAFRAS System Programs

The SAFRAS system consists of twelve COBOL programs. The decision to use COBOL as a programming language was based on two considerations. Due to the nature of much geological data, a programming language with good character manipulation capabilities was required. In addition, a programming language which could be used on a number of different computers necessitated the selection of a standard high-level language. In practice, it has been found that, despite claims to the contrary, not all COBOL compilers, even among different machines of the same manufacturer, are the same. However, the differences are fewer than is the case with many other compilers.

Figure 1. Overview of U.W.O. SAFRAS system.

SAFRAS SYSTEM → USER-GENERATED SYSTEM

FUNCTION	PROGRAM GENERATORS	SKELETON PROGRAMS
File Generation	SYSGEN	FILGEN, IOPACK
File Listing	SYSGEN	LISTER
File Editing	SYSGEN	EDITOR
Data Retrieval	RETGEN	
File Merging	MERGEN	MERGER
File Sorting	SRTGEN	SORTER
"Source" document generation	SORGEN	

DATA SPECIFICATIONS

GENERATED PROGRAMS	EXECUTION	OUTPUT
FILGEN	FILGEN IOPACK	SAFRAS Data File
IOPACK, LISTER	LISTER IOPACK	Printed listing of File Contents
EDITOR	EDITOR IOPACK	Edited Data File
RETREV	RETREV IOPACK EVALU8	Selected data in an output format specified by user
MERGER		"Updated" file, "Augmented" file, File as the sum of two or more files
SORTER		Sorted Data File
SORGEN		A Data-recording Document

Six of the SAFRAS programs are complete COBOL programs, and with the exception of EVALU8 (Figure 1), these are programs designed to generate COBOL source statements which are inserted into other "partial" or "incomplete" COBOL programs. It is for this reason that the following programs are called "Program Generators" (Figure 1).

SYSGEN – User system programs generator
RETGEN – Retrieval program generator
MERGEN – Merge program generator
SRTGEN – Sort program generator
SORGEN – "Source (data recording) document" program generator

There are six incomplete programs which have been termed Skeleton Programs. These are COBOL programs in which, for all intents, the data definition portion has been omitted. The SYSGEN program inserts one data definition into the Skeleton FILGEN, IOPACK, LISTER and EDITOR programs. Similarly, MERGEN and SRTGEN insert a data definition into the MERGER and SORTER Skeleton programs respectively. The RETGEN and SORGEN programs generate an entire COBOL program and therefore require no programs analogous to the Skeleton programs associated with the other Program Generator programs.

The result of the Generator Programs functions is the series of Generated Programs (Figure 1) which constitute a system based on a specific set of data. Within limits, any number of systems, based on different data sets, can be generated using the basic SAFRAS Program Generator and Skeleton programs.

SAFRAS Functions

When the SAFRAS system has been applied to the production of a "User-generated system", the following functions can be performed:

1. **File Generation.** The generated FILGEN program, in conjunction with the generated IOPACK program (i.e. the "input-output" routines for a SAFRAS System data file) will read the free-form format data, format the data and generate a formatted data file. In the process, errors such as "alphabetic characters in a numeric field", "too many fields in a record", "too many characters in a field ", etc., are detected and listed. The file can be "written" on magnetic tape or disc.

2. **File Listing.** The generated LISTER program in conjunction with the IOPACK program will produce a complete printed listing of the contents of a SAFRAS file. This listing includes both the values and the "value identifier" or "data item name".

3. **File Editing.** The generated EDITOR program linked with the IOPACK program allows a user to add, alter or delete the contents of a specific data field within a SAFRAS file. The user can also "create" an entire record which was not included in a specific unit of a file.

4. **Selective Retrieval.** On the basis of a specific set of data and a specific retrieval request, the RETGEN program generates a retrieval program called RETREV. This latter generated program, when linked with the generated IOPACK program and the retrieval condition evaluator program (EVALU8), searches the file and outputs the data which meet the retrieval parameters on punch cards, magnetic tape or magnetic disc in a format specified by the user in the retrieval request.

5. **File Merging.** The MERGEN program, on the basis of up to three sets of data definitions generated into the MERGER Skeleton program, results in a generated MERGER program which will perform one of four functions. These are:

Data Identifier		Field Size	Data Type#
*0101			
Deposit-Name	(a_1)	32	X
Latitude	(a_2)	2.5	N
Longitude	(a_3)	3.5	N
*0203			
Commodity	(b_1)	12	X
Grade	(b_2)	5.4	N
Grade-Units	(b_3)	12	X
*0301			
Primary-Reference	(c_1)	48	X
Author	(c_2)	24	A

# X = alphanumeric
A = alphabetic
N = numeric

Figure 2A. SAFRAS data-specifications.

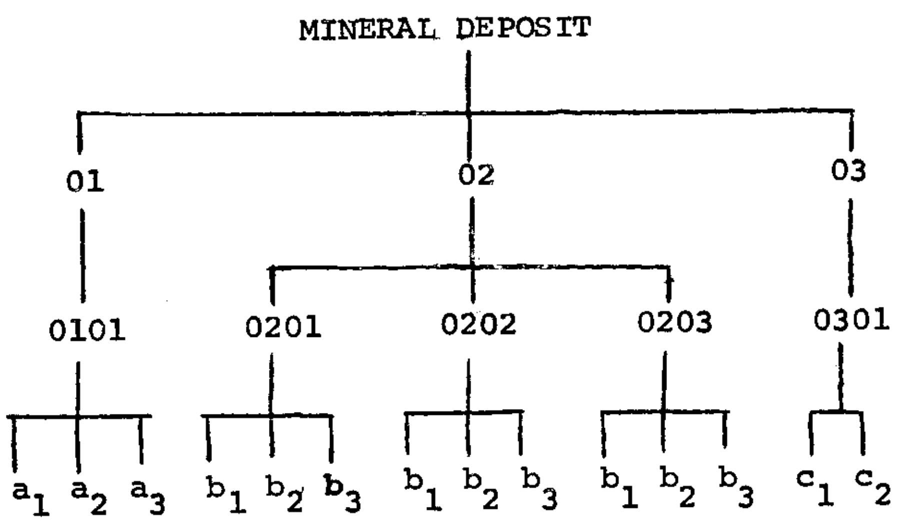


Figure 2B. Schematic representation of data structure.

1. **File Updating:** If, say, the basic unit of a file is a mineral deposit, the MERGEN program will allow the user to add additional mineral deposits to an exiting file which has the same data definition as the one included in the MERGEN program being used.

2. **Subfile Creation:** On the basis of two data definitions, the generated MERGER program will produce a file which although basically the same as an "original" file, has fewer items of data in each basic unit of the file.

3. **File Augmentation:** On the basis of three sets of data definitions generated into the MERGER program, provision can be made to add data to an existing file. The distinction between this function and the EDITOR funtion is that file augmentation involves the creation of "new" data fields which were not in the "original" file.

4. **File Merging:** Again, on the basis of three sets of data definitions generated into the MERGER program, any two (or more – two at a time) SAFRAS files having the same basic file unit can be combined into a single file.

6. **File Sorting.** In order to merge two (or more) sequential files, the file basic units must be in the same order. The generated SORTER program will sort one or more SAFRAS files into a sequence based on any number of sort parameters. For example, the deposits in a mineral deposit file could be sorted according to latitude within longitude within country.

7. **Source Document Generation.** A document, containing all the required control character information to generate a SAFRAS file from free-form format data input, can be produced in the form of a printed listing by the generated SORGEN program. This document is, of course, different for each set of data.

Data Specifications

The data specifications, which describe a specific hierarchical data structure, are the link between the SAFRAS System programs and the User-generated System programs (Figure 1). It is these data specifications which are translated by the Program Generator programs into COBOL data definitions. A simple set of data specifications, along with the structure they would represent are illustrated in Figures 2A and 2B respectively. The first two numbers which follow the asterisk at the head of each group of data item designators serves to identify that "set". The second two numbers dictate the maximum number of times a particular set can occur in each basic unit of the file, thus defining the hierarchy.

Summary and Conclusions

In order to provide a mechanism to support the SAFRAS system for potential users of the system, a copyright and a registered trade mark for "U.W.O. SAFRAS" were obtained and consigned to the University of Western Ontario. The University licenses the system for a nominal sum, in return for which the University commits itself to providing one year's support. Although expenses have consistently exceeded revenue, it is felt that the advantages of a supported system, together with the low aquisition cost, have been significant factors in encouraging ten organizations in Canada, the United States of America and Japan to license and use the SAFRAS system. What has been even more significant is that a continuing dialogue has developed with many of the system's users which is resulting in modifications to the system which should ultimately benefit all the users, present and future. This dialogue has been expedited by the Canadian Centre for Geoscience Data.

Selected References

Brisbin, W. C., and Ediger, N.M., Editors
1967 A national system for storage and retrieval of geological data in Canada: Nat. Adv. Comm. Res. Geol. Sci. (available from the Geological Survey of Canada, Ottawa), 175 p.

Burk, C. F., Jr.
1971 Computer-based geological data systems: an emerging basis for international communication: Proc. 8th World Petrol. Congress, v. 2, p. 327-335.

Hubaux, A.
1973 A new geological tool – the data: Earth Science Reviews, 9, p. 159-196.

Laznicka, P.
in press MANIFILE: the University of Manitoba computer-based file of World's nonferrous metallic deposits: Canadian Centre for Geoscience Data, Ottawa, Geological Survey of Canada Paper 75–20.

Sutterlin, P. G., and De Plancke, J.
1969 Development of a flexible computer-processable file for storage and retrieval of mineral deposits data: Proc. Symp. on Decision-Making in Mineral Expl. II, Feb. 1969, Univ. British Columbia, p. 11-42, 11 figs.

Sutterlin, P. G.
1971 The design of computer-processible information systems: in Decision-making in the Mineral Industry, Canadian Inst. Min. and Metallurgy, Spec. v. 12, p. 399-403.

Williams, G. D., Dickie, G. J., and Steiner, J.
1972 Computer storage and retrieval of geological data on coal deposits: Proc. First Geol. Conf. on Western Canadian Coal, Res. Council of Alberta Info. Series. Series no. 60, p. 73-84.

Etat actuel du système d'information mis en oeuvre pour la géochimie au Centre de Recherches Pétrographiques et Géochimiques (CRPG)

Ph. Grandclaude
Centre de Recherches Pétrographiques
et Géochimiques
15, Rue N.-D. des Pauvres
Case Officielle No 1
54 - Vandoeuvre-lès-Nancy
FRANCE

Introduction

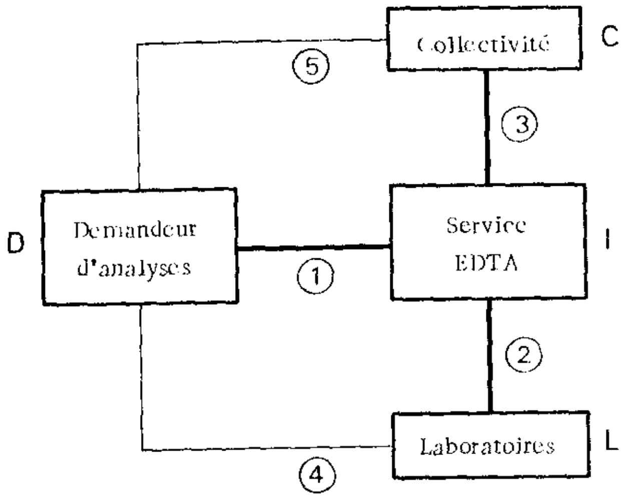
La mise en place au Centre de Recherches Pétrographiques et Géochimiques d'un quantomètre ARL a permis, depuis 1960, de produire chaque année plusieurs milliers d'analyses (majeurs et traces) de roches et de minéraux. Un tel volume de données, trop abondantes pour être communiquées au moyen des supports d'information traditionnels exigeait de nouvelles procédures et de nouveaux outils de stockage, traitement et communication de l'information.

En 1966, se tient à Nancy, sous l'égide de l'I.U.G.S. la première réunion "pour le stockage, la recherche documentaire et le traitement automatique de données géochimiques", qui rassemblait les représentants d'organismes confrontés sur le plan local à des problemes similaires et désireux, sur le plan international, de faciliter la communication des données qu'il convenait de joindre aux données géochimiques.

C'est à partir de cette liste que fut conçu un "système descripteur" des échantillons géochimiques, présenté sous la forme d'un "carnet opérationnel" utilisable pour la collecte des données de terrain et de laboratoire (localisation, conditions de prélèvement, environnement géologique local, âge, description pétrographique et minéralogique). Ce carnet est utilisé au CRPG depuis 1968 par les géologues demandeurs d'analyses. Jointes aux données géochimiques, les données correspondantes sont centralisées et stockées sur ordinateur. Ceci en vue de rendre accessible à une large collectivité, pour des recherches aussi diversifiées que possible, des données géochimiques créées au départ à des fins particulières.

Le Bilan en 1973

Plus de 30 000 analyses de roches et de minéraux ont été produites par le quantomètre, l'analyse standard fournissant actuellement 17 éléments dont 7 éléments de traces. Ces analyses sont effectuées dans le cadre de projets variés, depuis l'étude systématique des massifs granitiques jusqu'à des études géochimiques de séries sédimentaires.



——— Liaisons fonctionnelles

①	D → I	:	Transmission de la demande d'analyse et du carnet opérationnel
	I → D	:	Transmission des résultats d'analyse et retour du carnet opérationnel
	D → I	:	Demande et exécution de traitements sur données personnelles
②	I → L	:	Transmission de la demande d'analyse et d'un descriptif des échantillons
	L → I	:	Transmission des résultats d'analyse
③	C → I	:	Demande d'extraction de données et de traitement
	I → C	:	Transmission des résultats de traitement

——— *Autres liaisons*

④	D ↔ L	:	Détermination éventuelle d'un programme analytique spécifique
⑤	D ↔ C	:	Echange d'informations (publications, relations interpersonnelles)

Figure 1. Organisation de la collecte et du traitement des données géochimiques et des données adjointes.

Environ 14 000 analyses, accompagnées des informations collectées au moyen du carnet opérationnel ont été constituées en fichiers gérés sur ordinateur.

L'extraction à partir de cette "base de données" de fichiers de travail spécifique de projets déterminés et le traitement de programmes d'application conçus et réalisés par l'équipe d'informatique géologique du CRPG, a débuté cette année et fourni des résultats intéressants.

Description du Système

La figure 1 présente l'organisation adoptée: le service "Etudes Documentaires et Traitement Automatique de l'Information en Géochimie" (EDTA), placé en interface entre les demandeurs d'analyses et de traitement (de leurs données personnelles), les laboratoires d'analyses et les géochimistes ("collectivité") demandeurs d'informations (et de traitement) extraites de la base de données.

La figure 2 présente en son état actuel le système d'information et ses diverses fonctions:

- Collecte des données (bulletins d'analyses et carnets opérationnels)
- "Formalisation" des données: rédaction (format libre – langage proche de Signi) de bordereaux de perforation
- Saisie des données: cartes perforées
- Stockage des données sur bande magnétique (format fixe)
- Documentation des utilisateurs sur le contenu de la base de données (répertoires, cartes de prélèvements, statistiques)
- Sélection, extraction et "évaluation" (cf. documentation) de fichiers de travail à structure normalisée compatible avec les entrées dans les programmes d'application.
- Programmes d'application (adaptés ou en cours d'adaptation aux fichiers de travail)
    - . Analyse multivariable
    - . Calculs géochimiques et diagrammes
    - . Cartographie: cartes en courbes isovales, blocs isométriques, représentations en aires ou densités variables, représentation simultanée de deux variables.

Bibliographie

Govindaraju, K.
1973 New scheme of silicate analysis (16 major, minor and trace elements) based mainly on ion exchange dissolution and emmission spectrometric methods: Analusis, v. 2, no. 5, p. 367-376.

Grandclaude, Ph.
Contribution à la méthodologie d'un système d'information en Géologie: Application à la Géochimie. Mém. Sc. de la Terre, Nancy (à paraître).

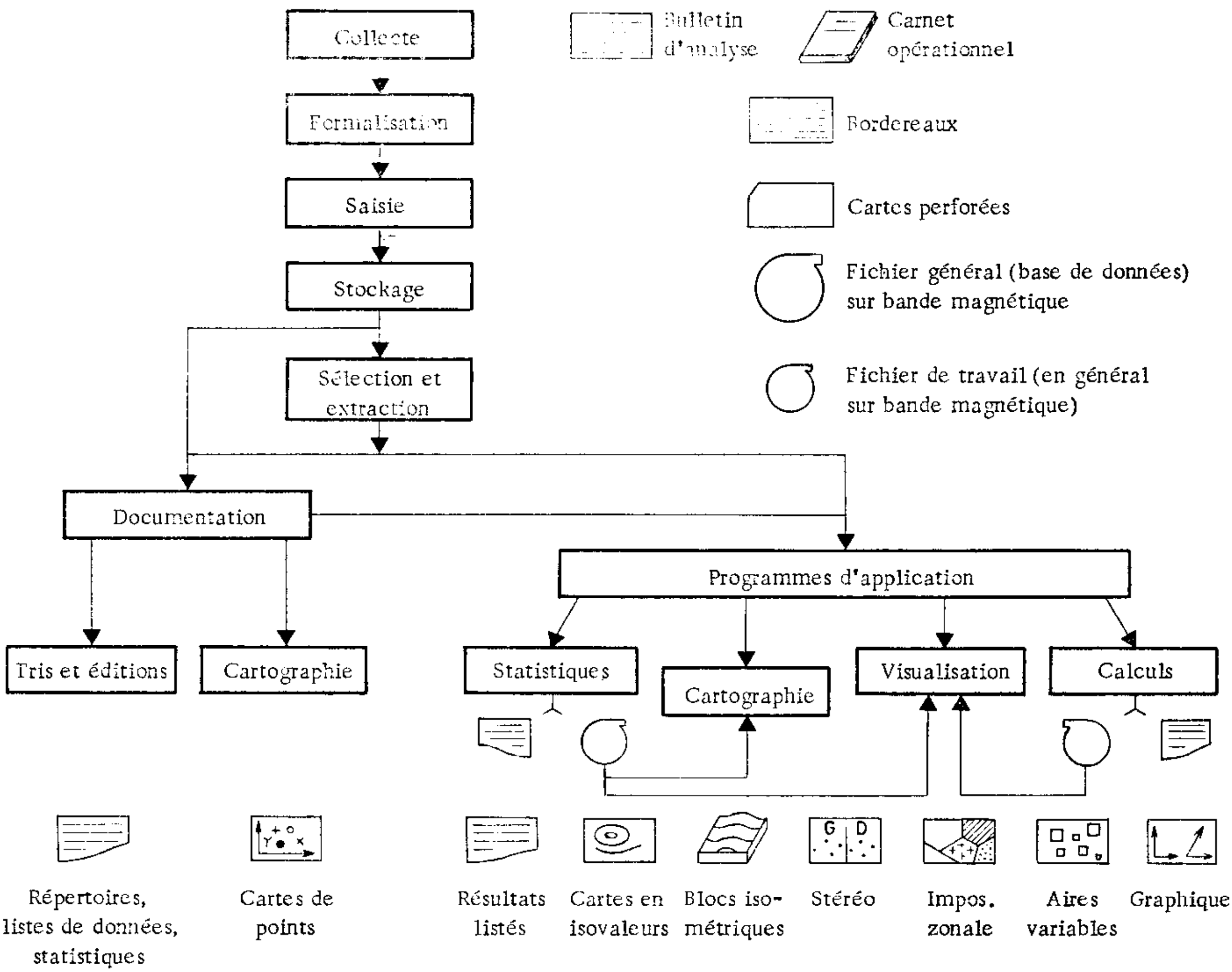


Figure 2. Organigramme général simplifié de la constitution et de l'exploitation de la base de données du C.R.P.G.

Isnard, P.
1970 Analyse statistique multivariable des données géochimiques par traitement automatique: Application au massif granitique de la Marche Orientale. Sc. de la Terre, Nancy, 15, no. 2, p. 103-143.

La Roche, H. (de)
1968 L'évolution des fichiers géochimiques et leur adaptation à la communication: CRPG, Nancy, doc. ronéo., 6 p.

La Roche, H. (de) et Grandclaude, Ph.
1968 Un système descripteur des échantillons géochimiques: Présentation du projet du Comité pour le Stockage, la Recherche Documentaire et le Traitement Automatique de Données Géochimiques, CRPG, Nancy, doc. ronéo, 9 p.

La Roche, H. (de)
1969 Enquête sur la présentation des données géochimiques: CRPG, Nancy, doc. ronéo. Fr. et Angl., 16 p.

La Roche, H. (de)
1970 Observations et données concernant les échantillons géochimiques: Notice d'utilisation du carnet opérationnel et exemples-type. CRPG, Nancy, doc. ronéo., 50 p.

La Roche, H. (de)
1973 Le service d'analyse de roches et de minéraux: Courrier de CNRS, no. 8, p. 36-43.

Leymarie, P.
1970 Contribution aux méthodes d'acquisition, de représentation et de traitement de l'information en Géologie. Sc. de la Terre, Nancy, mémoire no. 18, 170 p.

Mallet, J. L.
Méthodes et techniques de cartographie automatique numérique. Mém. Sc. de la Terre, Nancy, (à paraître).

Masset, J. M.
1973 Un système de visualisation des variations géographiques d'un paramètre géologique. Mém. Sc. de la Terre, Nancy.

Computer processing of cartographic data

L. W. Thorpe
Experimental Cartography Unit
Royal College of Art
6A Cromwell Place
London, S. W. 7
UNITED KINGDOM

The Experimental Cartography Unit (ECU) is basically an organization carrying out experimental work in developing, using and assessing cartographic editing systems.

The system presently being developed consists of approximately 14 programs which deal with data from the stage of digitization from base maps on a digitizing table, to the accurate replay of the processed and edited data on a high-resolution flat-bed plotter, the output from this device being suitable for printing purposes.

The following is a brief description of each of the programs:

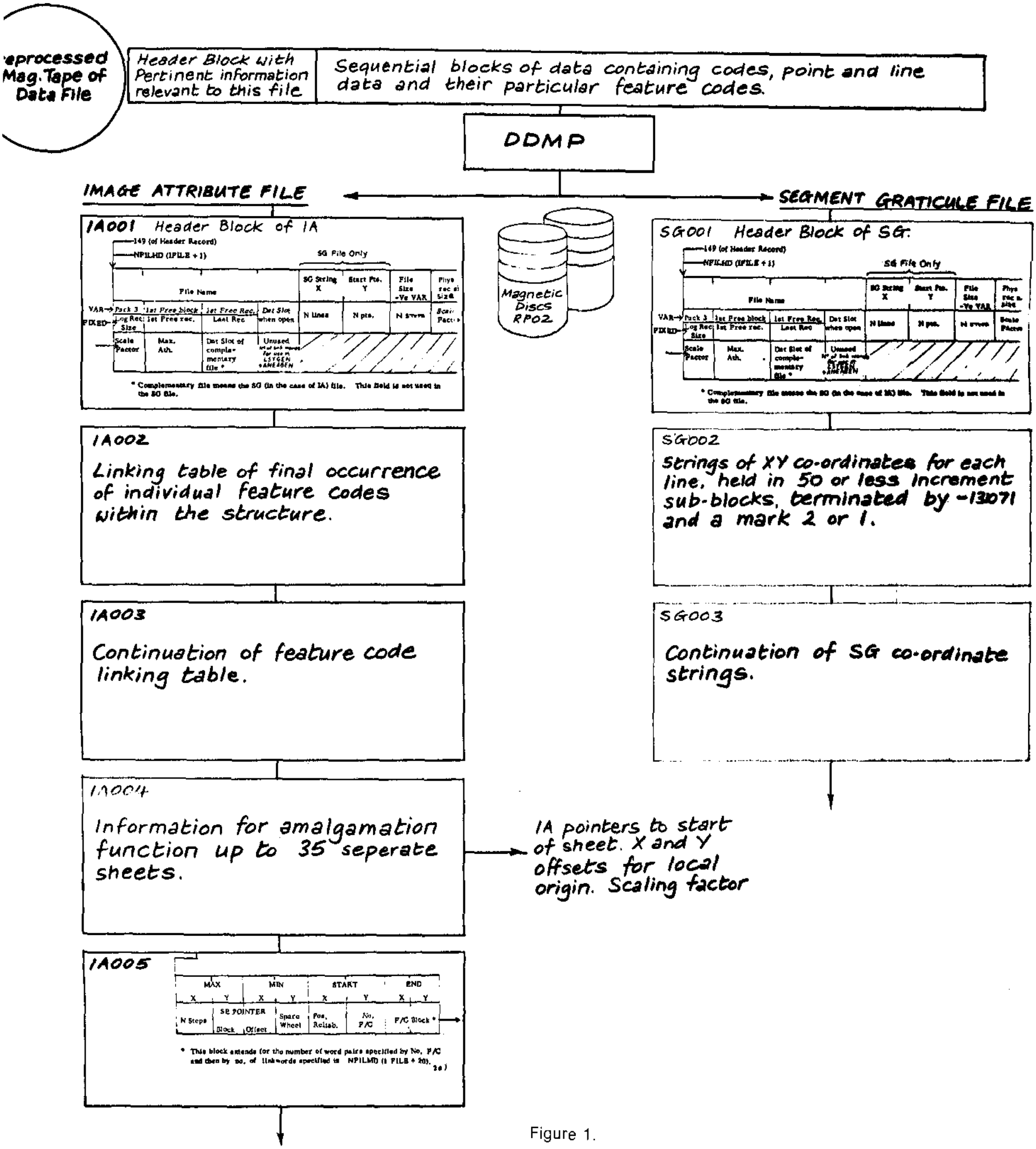
On-line Digitization

An interactive program run on a PDP 11 computer, via an interfaced D Mac digitizing table, which block-formats the coordinate data and associated feature code information of selected portions of a total map to be digitized onto a DEC tape. Illegal sequences of operations are prevented. Digitizing mistakes recognized by the operator at the time of digitizing can be corrected or edited out. Maximum and minimum values of feature codes and number of feature codes associated with each map element are checked and diagnostic messages given in cases of errors.

The DEC tape is subsequently processed via a PDP 9 or -15 onto a high-density magnetic tape. This tape has a header block containing geographical, mnemonic, projectional and scaling information about the map that follows. The tape may contain more than one map sheet and each sheet is prefaced by such a header block. The data on each sheet is stored as X-Y coordinates, either in the form of a line or a point.

Totals and Syntax Checks

This program checks each of the map sheets on a magnetic tape for the presence of digitized corner points and internal consistency of character strings. Any errors present on the tape produce



Preprocessed Mag.Tape of Data File
Header Block with Pertinent information relevant to this file
Sequential blocks of data containing codes, point and line data and their particular feature codes.
DDMP
IMAGE ATTRIBUTE FILE
SEGMENT GRATICULE FILE
Magnetic Discs RPO2
IA001 Header Block of IA
SG001 Header Block of SG.
SG File Only
File Name
IA002
Linking table of final occurrence of individual feature codes within the structure.
SG002
Strings of XY co-ordinates for each line, held in 50 or less increment sub-blocks, terminated by -131071 and a mark 2 or 1.
IA003
Continuation of feature code linking table.
SG003
Continuation of SG co-ordinate strings.
IA004
Information for amalgamation function up to 35 seperate sheets.
IA pointers to start of sheet. X and Y offsets for local origin. Scaling factor
IA005
MAX
MIN
START
END

Figure 1.

diagnostic output and where possible are automatically corrected. The corrected tape has the header updated with the number of points, increments,lines and maximum number of feature codes associated with any line or point on this sheet. This information is used in the preceding program to define random access areas on fixed or moving head discs. Re-ordering of the sheets on the tape is also available.

Disc Formating from Magnetic Tape

From the magnetic tape produced by the above programs, a random access data structure is developed on a disc. The form of the structure is demonstrated in Figure 1.

Line Start/End and Code Verification

Each map sheet is generally digitized by one operator and then just the starts and ends of lines with their codes are redigitized by a second operator. These two files are processed by the previous programs to the stage where they are automatically compared to produce diagnostic output about any code differences, and line or point omissions, of one file compared with the other. This technique produces a good check that all the lines to be digitized have been, and that their codes are what they should be.

Disc Editing

Online editing of the data in the structure to delete lines, points and change feature codes on elements of the map recognized by program 4 to have possible errors.

Linking of Feature Codes

This program has the facility to give a listing of all the elements of the map, with their starts and end points and their feature codes, plus the facility to produce a table, in blocks I and II of the Image Attribute file shown in Figure 1, of feature codes, and a pointer to the end of a 'LAST IN FIRST OUT' list. This technique allows efficient retrieval of elements of the map containing particular feature codes or combinations of feature codes.

Interactive Graphics and Editing

This program displays the map on a CRT screen. It allows the display of the total map or selected regions at different scale: elements with selected feature codes or combinations of feature codes can be displayed at the selected scale, in various pecking and hatching patterns. It is possible to directly delete lines or points, or alter their feature codes. Lines can be split, joined together and even new lines directly inserted. Similarly points may be inserted. Various grids can be superimposed on the screen representing the geographical locations, digitizing scale, or even screen coordinate system.

Amalgamation

Sheets that have been checked and interactively edited can be amalgamated into larger files. This can be done in a vertical or horizontal sense, i.e., a weather map of an area can be amalgamated with a geological or political boundary map or two adjacent geological maps may be amalgamated together. Any number of such sheets can be added together in one operation of the program or subsequent operations.

Overlaying of Two or More Vertically Amalgamated Sheets

The output from the previous program will in general contain lines that intersect. This can be envisaged using the weather and political boundary map quoted earlier. This program locates all line intersections and updates the structure to contain such lines split at these points of intersection.

Thus the output of this program will contain no line that intersects with any other line.

Logical Combination of Feature Codes

This program retrieves from the data structure two sets of elements with specified feature codes. It then produces a sequence of pointers on the particular IA records that satisfy the BOOLEAN OPERATION specified for these codes.

The available operations include AND, ORR, NOT, and OXR. Having once produced a combination in this way one may combine this set with others as described above. These sets may be retrieved in any of the relevant programs making up this system for further checking, or further processing by programs described later.

Closed Area Recognition and Retrieval

Sets of lines that have been developed by the previous program or a set of lines with a particular feature code may form a number of closed areas. A program exists to automatically recognize such closed loops and update the structure with AREA records that describe the area in a similar manner as lines are described.

These areas can be retrieved and displayed, areas can be measured, and a point can be associated with the areas within the closed loop, i.e. the point in polygon.

Unamalgamation of Data Sets

When a number of vertical and horizontal amalgamations have been performed and overlayed, the data set becomes quite large and possibly a little cumbersome for some applications. Similarly when two very large files have been amalgamated and a edge-match problem exists, a small file in the region of the edge-match error is much easier to handle than the total file. This program allows a selected rectangular area of a file to be created on a separate structure for further editing work before being reamalgamated into the final file.

Unformating of Disc onto Magnetic Tape

For plotting purposes and long-term storage the total disc structures can be dumped onto tape or selected elements may be dumped. The sets that can be retrieved are the different feature code sets, the feature code combinations, and the area sets.

High-Accuracy Plotter
The tape produced by the previous program can be plotted on the high accuracy plotter by a 'universal plotting package' and the output is quite suitable for printing and publication purposes.

A number of other programs exist that have not been included in the above list, which are used for data reduction in the sense of salient point production along lines and then the recreation of the original points from these salient points by splining. Two programs exist for very close examination of the data on the disc in a block by block examination mode, and block by block editing mode.

Particular requirements for the Dutch WIA Earth Science Information System

T. de Heer
IWIS-TNO
P. O. Box 297
The Hague
NETHERLANDS

S. W. Bie
WIA, Stiboka
P. O. Box 98
Wageningen
NETHERLANDS

Introduction

The Geological Survey of the Netherlands and Netherlands Soil Survey Institute, in recognition of the considerable similarity of their tasks, have joined forces in development of an automated information system for earth science data (WIA). Preliminary discussions started in 1970, with a construction phase commencing in the autumn of 1972. We envisage the framework of the system operating by the end of 1975.

The project is seen by the Dutch government as an experiment in the more general problems of environmental and planning information systems. It is therefore our objective to make a generalized system, where ad hoc data structures and retrieval systems have no place. A close review of the data collected by the two surveys similarly illustrate the desirability of a generalized system.

Input Data Requirements

We find that our main objection to nearly all existing data base management systems is that they are unable to handle satisfactorily "free-language" information with variable record lengths and structures. So here we will concentrate on this problem.

The Geological Survey has at present about 125 000 logs of shallow boreholes illustrated in Figure 1, increasing by 10 000 per year. In addition, they have some 45 000 deep borehole records of similar type. The Survey requires most of these records to be available in an automated system. The recording and translation into numeric form of fixed format of all these data requires unacceptable investments of manpower, time and money. Most critically there are not the kind of skilled personnel available and willing to do it. The alternative is to build a system which will

No 368-24-1 Profiel 1,47 m + NAP

Diepte	Grondsoort	Omschrijving	Leem	Lu	M 50	Ca
0,00 - 0,20	klei	met scrobicularia		27		4
0,20 - 0,40	klei	veel scrobicularia		15		4
0,40 - 1,00	zand	licht, gelaagd, met roestvlekken		4	100	4

Figure 1.

accept and process records of variable length and structures, and "natural language contents" with data preparation performed by not so highly skilled staff. Recent developments have made available search techniques and search strategies which operate on fully free-format natural language files, finding all answers on queries in the form of boolean expressions. The operands of the boolean questions are substring specifications, which the system tries to find as substrings of the natural language contents of the file-records.

The system must have a data entry facility which accepts input data in "free format" (as well as in the more normal "fixed format") and to check the input data with respect to a defined "input data description" (IDDL). The "input data description language" defines "free-format input data structures" as hierarchies of attribute/value-pairs.

An example of the possible shape of an input record is given below (based on Figure 1):

```
$ PROFILE $ 368-24-1    $ NAP $ + 1.47
$ LAYER $ 0.00/0.20     $ SOIL $ CLAY
                        $ DESCRIPTION $ WITH SCROBICULARIA
                        $ LU $ 27
                        $ CA $ 4
$ LAYER $ 0.20/0.40     $ SOIL $ CLAY
                        $ DESCRIPTION $ MUCH SCROBICULARIA
                        $ LU $ 15
                        $ CA $ 4
$ LAYER $ 0.40/1.00     $ SOIL $ SAND
                        $ DESCRIPTION $ BRIGHT, LAYERED, WITH
                          RUSTSPOTS
                        $ LU $ 4
                        $ M50 $ 100
                        $ CA $ 4
```

An input-stream like this could be defined by means of the following IDDL statements:

```
NAME = "PROFILE"; VALUESET="I3-2-1"; GROUP; FORMAT=FREE; MANDATORY
NAME = "NAP"; SUB=(PROFILE); VALUESET="N(-10/500)"; FORMAT=FREE;
       MANDATORY; SINGLE
NAME = "LAYER"; SUB=(PROFILE); VALUESET="N1.2(0/12)$(1)$N1.2(0/12)";
       GROUP; FORMAT=FREE; MANDATORY
NAME = "SOIL"; SUB=(LAYER(PROFILE)); VALUESET=(CLAY,SAND); FORMAT=FREE;
       MANDATORY; SINGLE
NAME = "DISCRIPTION"; SUB=(LAYER(PROFILE)); FORMAT=FREE; SINGLE
NAME = "LU"; SUB=(LAYER(PROFILE)); VALUESET="I3"; FORMAT=FREE; SINGLE
NAME = "M50";          idem
NAME = "CA";           idem
```

By this definition-form one can specify for each attribute-name:

- the set of values allowed for that attribute (VALUESET = valuset specifications with structure code)
- the hierarchical position of the attribute with respect to higher-level attribute (SUB=)
- the requirements for presence of the attribute: some attribute must be present (MANDATORY) whereas others just may be present or not, which can be dependent of the presence of higher-level attributes which an attribute is subordinate to.

The input stream can be checked by the system on agreement with these specifications.
The record type involved needs a so-called record-directory, a particular part of the record which tells the system about the location and field-length of each attribute it contains. A much applied type of record which agrees with these requirements is the MARC II TYPE record format.

Information-retrieval Requirements

The system should accept and answer questions put to the systems by the users as follows:

- queries should be allowed in the form of boolean expressions
- the operands of the boolean expression should specify a simple condition with respect to the value of a single attrubute
- the simple conditions should, in addition to the normal relational expressions like NAP.< 1,50,LU < 20 or DESCRIPTION = "WITH SCROBICULARIA", provide the possibility of "substring specifications", such as DESCRIPTION.SU."SCROBIC".

Data-base Requirements

The data-base system should be able to handle four main types of data sets:

1. Point data sets
2. Line data sets
3. Area data sets
4. Descriptional data sets, the attributes of which are independent of geographical coordinates.

A second requirement is ability to perform efficient retrieval on the types of queries described earlier, and on normal numeric and alphanumeric fixed and free-format records.

A third reqirement is that the data-base system should be able to handle relations between the system-objects (records) of the data sets such as:

- relations between point-records and line-segments telling the system that a geolographical point lies to the right or to the left of a given line segment with respect to the direction of its point-chain representation
- relations between point-records and areas telling the system of which areas a given geographical point is an inner point
- relations between descriptional-attributes and lines and areas telling the system to which lines and/or areas the descriptions apply.

Fourthly the data-base system must have a two-directional interface with a map drawing system which may consist of small computers and drawing machines, display units and map digitizers and to be available for inter-active work.

Fifthly, we desire that the DBMS have a high degree of portability.

Conclusions

Our data handling involves the scanning of large and varied data sets. We do not have manpower available for standardization and coding of geological borelogs. We need a system that can accept and process natural language of variable lengths and structure and offer powerful input checking procedures. To these specifications we wish to construct or modify an earth science information system.

Land-use planning aided by computer cellular modelling/mapping system to combine remote sensing, natural resources, social, economic, and cadastral data

Harry W. Smedes
United States Geological Survey
Building 25, Federal Center
Denver, Colorado 80225
UNITED STATES OF AMERICA

Summary

Land-use planning, land-management, and evaluations of the environmental impact of specific changes in land use require a consideration of the total environment. Included are map, point, and tabular data of such varied attributes as land-cover or terrain types; surface and subsurface natural physical features such as slope, landform, lithology, thickness and nature of surficial deposits, surface and subsurface hydrology, vegetation, soils, wildlife habitat, and rangeland quality; ecology; social-economic features such as income, ethnic concentrations, and available labor skills; and point features such as locations of key facilities.

Some of these types of data can be acquired by remote sensing, and others must be acquired from conventional sources. In order to consider the total environment, we must accelerate the acquisition of pertinent data. However, remote sensing and each of the many separate and necessary disciplines whose sound basic data are essential to this total land-use decision-making have largely developed independently. Consequently, many of the available data are peer-oriented and not readily comprehensible to users outside each discipline. Therefore, there is an urgent need to devise a system of combining and manipulating these new and existing data, and to present them in understandable form.

This paper presents a means for combining and analyzing these diverse types of environmental data in a common format by way of a cellular composite computer mapping system. In experiments in Jefferson County, Colorado, and in southeastern Missouri, this system is effectively being used as an aid in working toward the timely solution of such commonly occurring land-use problems as selecting optimum sites for: acquisition of open-space and greenbelts, sanitary land-fill, sewage treatment lagoons, septic tanks, and housing developments.

Each attribute that is considered important in this land selection procedure (i.e., geology, soils, etc.) is mapped over the entire area. Each resulting conventional map is cellularized to produce a cellular matrix map. The values ascribed to each cell represent the characteristic of the selected attribute for that particular rectangular land area.

Cellular map matrices can then be quantitatively weighted, manipulated and combined according to land-use strategies determined by the local planners. Alternative solutions can be tested and evaluated. Later, using any one map as a dependent variable, an investigator can make a multiple regression analysis of the remaining independent variables (maps) to determine the possible predictable relations between maps, for example determining the relation of health to environmental factors. In addition, sensitivity analysis can be performed, that is, an analysis of the degree to which the result reflects changes in the weights assigned to a specific attribute. This analysis can give insights into the minimum number or minimum accuracy of attributes needed to make the optimum decision.

Especially important is that the County and State users have been actively involved in the design, evaluation, and evolution of these tests from the start. In Jefferson County, Colorado, users plan to take over the project at the completion of the research phase.

Whereas, each aspect was formerly considered separately, the present system enables them to be combined and manipulated for analysis; results have been synergistic.

This system involves the use of cellular data in digital format. Developments in remote sensing enable us to use computer (programmed with multispectral cellular digital data from sensors in satellites and aircraft) to map such terrain classes as water, vegetation, and soils, as well as changes in land cover. I envisage that these thematic maps can soon be linked as an important direct input into the composite computer mapping system.

Acknowledgments

These studies are sponsored by the U. S. Department of the Interior Resource and Land Information (RALI) Program in cooperation with the Missouri Geological Survey, the Federation of Rocky Mountain States, Jefferson County Open Space Advisory Committee and County Planning Department, the Colorado School of Mines, and the citizen's group Plan Jeffco.

Of the many people who contributed to this study, I wish especially to thank Douglas Mutter, George Nez, Larry Salmen, John Reed, and the late Edwin Lutzen.

Appendix

The particular cellular mapping system used was the Composite Computer Mapping (CCM) system developed by the University of Utah for the Economic Development Administration of the Department of Commerce. The original programs were written for UNIVAC 1108 computers. They have been rewritten for use with CDC 6400 series computers and are presently being rewritten for IBM and for PDP- 10 computers.

Reference

Hachman, F.C., Bigler, C. and Weaver, R.
1971 Report on the implementation of composite computer mapping: Bureau of Economic and Business Research, Centre for Economic and Community Development, University of Utah, Salt Lake City, Utah, U.S.A., 227 p.

Advantages of using a generalized system to manage geological data

G. D. Williams
Department of Geology
University of Alberta
Edmonton, Alberta T6G 2E1
CANADA

The terms 'data' and 'information' are sometimes used interchangeably, and sometimes attempts are made to separate objective, reproducible facts ('data') and the conclusions drawn from them ('information'). It can be shown however, that 'data' and 'information' are parts of a hierarchy of knowledge in which any element may be either 'data' or 'information' (Figure 1). For example, a surveyor uses 'data' in the form of instrument readings, angles and lengths, and produces 'information' in the form of point locations and elevations. A geologist may use, as 'data', this information, in conjunction with information produced by chemists, petrographers and others, process them and produce 'information' pertaining to the size, shape, type and origin of a mineral deposit, for example. This geological information, in turn may become 'data' for subsequent processing and ultimate production of more new 'information'. The distinction between 'data' and 'information' thus depends entirely upon our point of view, and the degree of reproducibility (or objectivity) of any 'data' depends upon their position in the hierarchy.

Many geological data can only be recorded with a low degree of precision, and in a very large number of cases, particularly in field geology, it is easier and more convenient to record only the conclusions drawn and inferences made, rather than the observations which led to the conclusions and inferences. In this regard, the field geologist usually functions as an interactive analog computer, interpreting observations as they are made and drawing conclusions from them, based upon his training, experience, personal bias and the influence of local conditions under which he is operating. In many cases the observations made and stimuli received by the geologist do not lend themselves to being recorded as data; consequently only the conclusions are recorded, often in map or other graphic form.

Although this procedure is undoubtedly effective in many cases, it is also extremely dangerous, because it may lead to neglecting even those observations which can be recorded for future study. The result is that at some later time, it may not be possible to verify the validity of the conclusions, perhaps in the light of new data or changes in geological philosophy. Even more crucial is the possibility that a critical item of data may not have been recorded, either through oversight or because it was so obvious at the time the observations were being made.

The use of computers to manage geological field data has focussed attention on the need to be systematic in the collection of the data. It has also become apparent that both factual observations, and the conclusions based on them should be recorded and distinguised as clearly as possible.

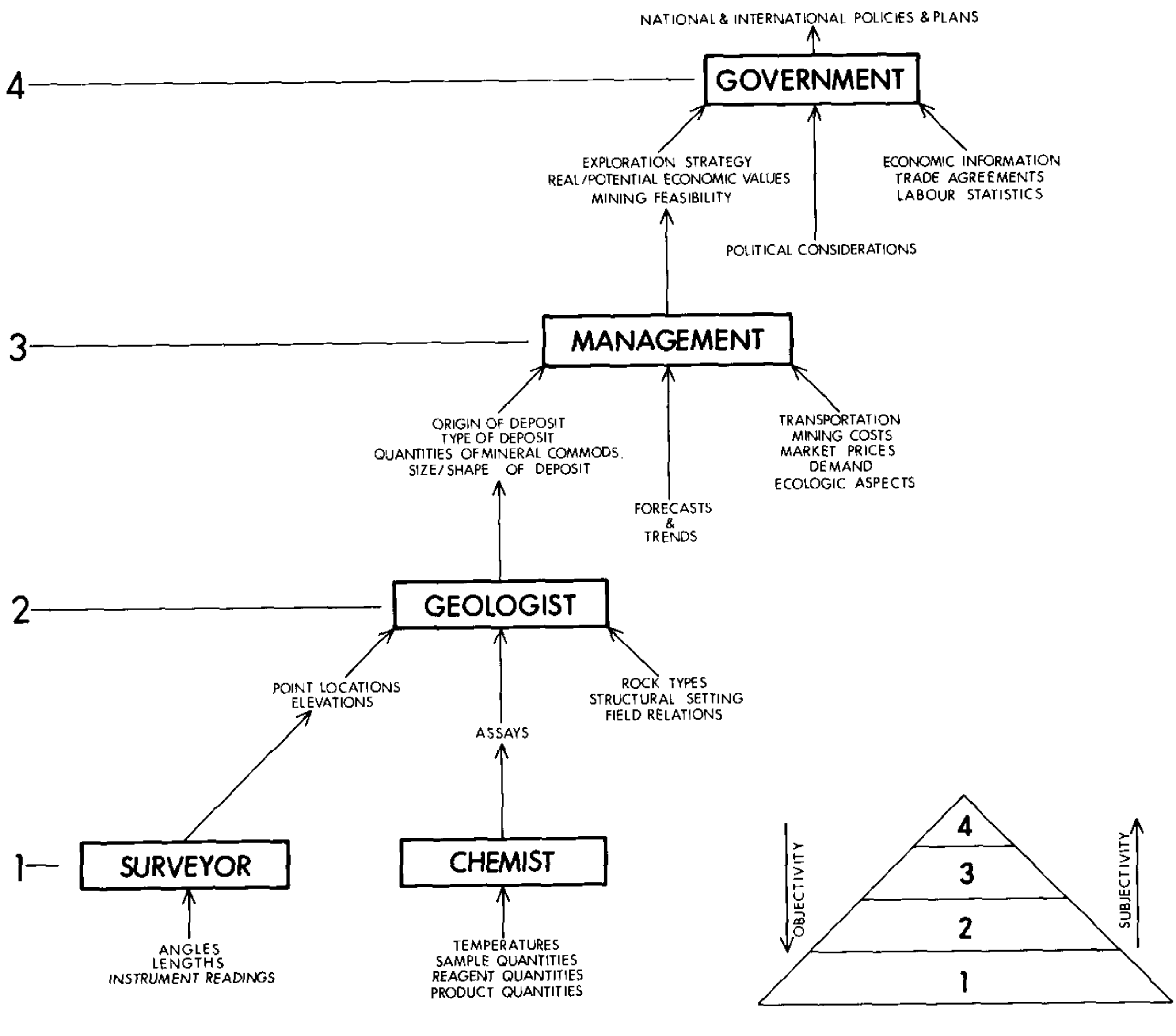


Figure 1. Hierarchy of "data" and "information".

Only a finite, small number of operations can be performed on data:

1. Collection
2. Storage
3. Updating
4. Rearrangement
5. Retrieval
6. Display
7. Analysis

Once data are collected, using whatever techniques are necessary, they may be subjected to any of these operations in any sequence. For example, data may be displayed (as lists, charts, maps, etc.) before or after updating, rearrangement, or analysis; errors are often detected by analyzing or displaying the data or analytical results; rearrangement, retrieval and display of selected items may indicate what new types of data should be collected, and so on. Where the quantity of data is relatively small, it is possible, and often convenient to go directly from collection to display and analysis without consciously considering the intervening operations. Indeed, many data are collected specifically to fit predetermined display or analytical schemes. Much of the collection, storage, display and analysis of geological field data for example, typifies this situation. This procedure becomes inefficient or unworkable, however, if:

1. The quantity of data becomes very large
2. The data are collected, even by a single individual, over a considerable period of time, or
3. The data are collected by several individuals in succession over a long period of time, or simultaneously over a large geographic area or from diverse sources.

To forestall these problems, a systematic formalized procedure must be used to handle the data collected. This involves building an organized structure for the data (a data file) and emphasizing the four management functions listed above (storage, updating, rearrangement and retrieval). Unfortunately, many geologists seem to believe that the way data are collected in the field (codes, input documents used, etc.) determines the way in which they can be processed, or that the kind of storage and retrieval (management) system determines how the data must be collected in the field. Although it may be convenient to record data in formats or on special input sheets which will facilitate subsequent computer processing, the processes and procedures involved in data collection are for the most part entirely separate from those involved in data management, and linking these two aspects merely imposes additional hardships and constraints on the field geologist. Whereas geological field data are undoubtedly unique in many aspects, the problems of managing them are not.

Data files may be managed either by constructing an ad hoc system tailored specifically to the data in the file and the perceived or anticipated uses to which they will be put, or by making use of a generalized management system designed to handle a wide range of data files and types.

An ad hoc system can be very efficient in its use of computer time, but it normally requires the user to have a considerable amount of computer knowledge and experience, or that he have access to a programmer or systems analyst who is knowledgeable in the user's speciality. The expense associated with anything more than the very simplest ad hoc system is usually quite large, because each operation performed on the file, be it entry of new data, retrieving and sorting or correcting (updating) data, requires the writing of a separate program or programs, and programming costs are high. Moreover, the builder usually finds that after a file has been partially constructed, he wished to change it in some way - add or delete data items, use different formats for some items, split items into two or more, etc., - and for an ad hoc system, this almost invariably requires complete redesign and reprogramming, with attendant expense and loss of time, effort and sometimes information. Unfortunately, there is every likelihood that the builder will want to make further changes when the 'new' file is partially constructed, and this procedure may be repeated several times, or the builder may decide to proceed using a system which is less than adequate. Expenses are likely to be high, regardless of the procedure adopted.

Generalized data management systems have been defined as "....user-oriented, generalized software packages designed for the maintenance of and selective retrieval from...files of various types and structures. The primary goal of these systems is the ability to manipulate data files, performing functions related to file definition, file creation, file interrogation and file maintenance." (Dolan et al., 1970). Key phrases in the definition include:

1. "user-oriented" - the user should be able to manage his own data without having to become a computer expert first, or without continuing assistance from professional programmers.
2. "files of various types and structures" - although most generalized systems have been designed for business applications, they can often be used equally well for scientific purposes; a few have been designed specifically for scientific uses.
3. "functions related to file definition, file creation, file interrogation and file maintenance" – a good system will assist the builder with all aspects of data management, from defining and creating a file to selective extraction of data items from it. Not all systems perform all functions with equal success; some systems may have good retrieval capabilities but may be woefully weak in their file-building or updating functions, or vice versa. For scientific applications, a system should be able to retrieve data in forms which can be used directly as input to analytical or display programs and packages. A good system is well balanced.

At the present time, there are several hundred systems available which purport to be "generalized'. Some are distributed by hardware manufacturers and some are marketed by companies specializing in software. The latter often have a fairly broad compatibility with different types of computer hardware, in contrast to the former.

The cost of acquiring a generalized data management system varies as widely as the range of services provided, from about $3000 to $50 000, and the most expensive is not always the best for a particular application. Obviously the amount of anticipated use must be able to justify the cost of acquisition, which on first inspection may seem to be inordinately high. However, when costs are being considered, it must be remembered that many thousands of dollars may have been spent to collect the data, and that many thousands more can very quickly be spent for salaries of programmers, analysts and geologists in trying to make a poorly-designed or otherwise recalcitrant system operate satisfactorily, not to mention the indirect costs of frustration created and time lost in the process.

Finally, on the subject of costs, it is totally unrealistic to believe that the use of any kind of computer data management system will lower the absolute costs related to using any collection of data. However, although absolute costs will normally increase, an effective system will increase cost efficiency, lower cost per unit of operations, and provide the opportunity to undertake tasks which otherwise would have been impossible to contemplate.

Selected References

Boehm, B. W.
1973 Software and its impact: a quantitative assessment; Datamation, May, 1973, p. 48-59.

Brisbin, W. C., and Ediger, N. M., Editors
1967 A national system for storage and retrieval of geological data in Canada; National Advisory Committee on Research in the Geological Sciences, Canada, Geological Survey, 175 p.

Dolan, F. T., Dust, N. H., and Woodcock, J. R.
1970 Evaluation of generalized data management systems; Internal Report, University of Calgary, Calgary, Alberta.

Madill, J. P., and Kuss, J.
1971 Study and recommendations for acquisition of generalized data management systems; Internal Report, Simon Fraser University, Burnaby, British Columbia.

Moody, D. W.
1972 Applications of generalized information systems to the storage and retrieval of geological and geographic information; Proc. 24th Int. Geol. Congress, Section 16, p. 161-168.

Roddick, J. A., and Hutchison, W. W.
1972 A computer-based system for geological field data on the Coast Mountains project, British Columbia, Canada; Proc. 24th Int. Geol. Congress, Section 16, p. 36-46.

Williams, G. D., Dickie, G. J., Gordon, T. M., Groen, H. A., Kelly, A. M.
1972 Future development of the SAFRAS generalized data base management system; Internal Report, Canadian Centre for Geoscience Data, Canada, Geological Survey, Ottawa.

Geologic Retrieval and Storage Program (GRASP)

Joseph Moses Botbol and Roger W. Bowen
United States Geological Survey
National Centre Mail Stop 904
12201 Sunrise Valley Drive
Reston, Virginia 22092
UNITED STATES OF AMERICA

The U. S. Geological Survey developed a Geologic Retrieval and Storage Program (GRASP) in response to an immediate need for interactive oil and gas pool data processing. The design philosophy of GRASP was largely governed by four parameters: complete system portability, immediate data access, simplicity of operation, and geoscience orientation. There are 400 variables for each pool in the oil and gas data file. In addition, there is a graphics file containing the boundary vectors for each pool. The files are linked via a unique identification number. Tabular data are input to GRASP via a twin magnetic tape cassette terminal or, if the data are computerized, by reformatting. Graphics data are input via a cathode ray tube (CRT) and drum-scanner technique. Output from the system is primarily via CRT, although future plans call for the use of other peripheral output devices that can be addressed from the CRT.

Introduction

About two years ago, the U. S. Geological Survey determined that batch storage and retrieval of tabular data were inadequate to meet programme demands. Furthermore, as geologists traditionally use maps for graphical representation of their data, any projected storage and retrieval system would have to have the capability of generating maps.

The U. S. Geological Survey, the American Association of Petroleum Geologists, and the University of Oklahoma, working on a cooperative oil and gas pool data bank project, evaluated many 'in-house' and commercially available information systems and decided that modification of any existing system would be both conceptually and mechanically too confining. The most efficient route, that of developing a Geologic Retrieval and Storage System (GRASP) that had the ability to manipulate both tabular and graphics data, was adopted by the U. S. Geological Survey.

The design philosophy of the GRASP system was governed by the following considerations:

1. Complete portability. The system can be installed on any system that has a standard ANSI FORTRAN IV compiler.
2. Immediate data access. The system was designed to be implemented in a time-sharing mode. This, of course, does not preclude its operation in batch mode.

Table 1. Tabular information categories of the oil and gas pool data file.

Identity
Vitae
Geology
Engineering
Reservoir rock properties
Development
Reservoir fluid properties and saturation
Production
Reservoir pressure and producing ratios
Hydrocarbon volume and primary rocovery
Secondary recovery
Chemical data: crude
Chemical data: natural gas
Chemical data: water
Discovery well data
Deepest well data
General comments

3. Simplicity of operation. There are 11 commands in the GRASP command language. Graphics are implemented by using a 'menu' technique.
4. Geoscience orientation. File structure and manipulation, graphics implementation, and general retrieval operations were all designed to be used in a way that a geologist might ask questions of a data bank.

At the present time the tabular storage and retrieval part of GRASP is fully operational. Various aspects of the graphics part of the system have already been successfully tested, and it is expected that the graphics will be fully operational by autumn 1974.

Oil and Gas Pool File Description

The following is a brief description of both the tabular data and the graphics data which comprise the oil and gas pool data file and upon which GRASP is presently functioning.

Tabular data. The entire data set is composed of about 400 variables for each pool. Because of the brevity of this report, only information category headings are listed in Table 1.

Graphics data. The graphics data file is composed of the boundary vectors that define the outlines of each pool in the system. Linkage to the tabular part of the system is by a unique pool identification number and by maximum and minimum northing and easting.

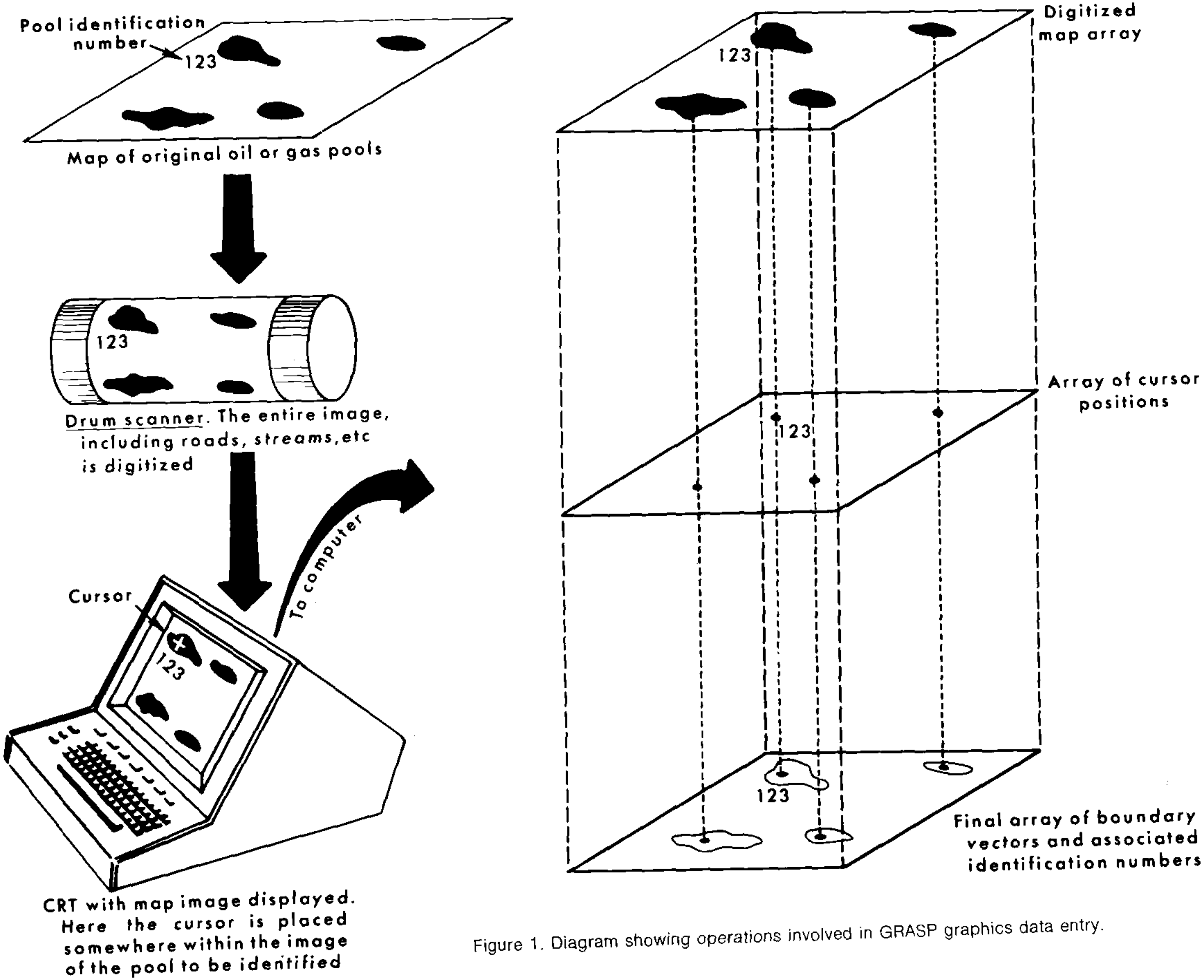


Figure 1. Diagram showing operations involved in GRASP graphics data entry.

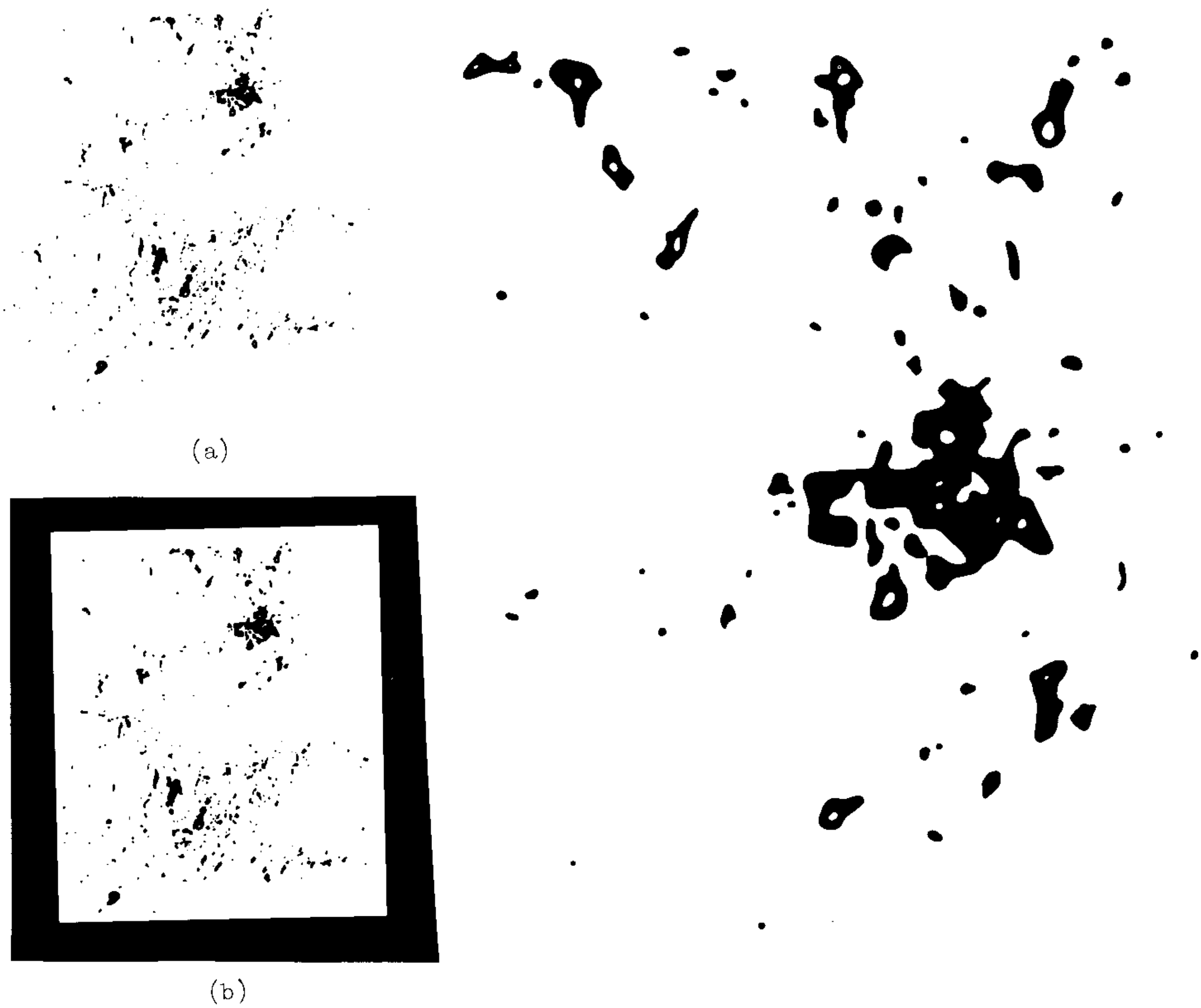


Figure 2. Comparison of original map photo reduction with scanned images.
(a) original map photo; (b) direct print of a 25u scanned image of (a); 2-step photo enlargement of the upper right quarter of (b) to a scale of 1:500 000 (the scale of the original map).

Data Input

Both the tabular and the graphics data each have their own distinct method of entry into GRASP. Tabular data occur in two types: data that pre-exist in another computer system and data collected especially for GRASP. Graphics data are entered directly from maps.

Tabular data input. Conventional reformatting programs are used to convert existing computerized data to GRASP. New data are entered to GRASP via a programmable terminal equipped with twin magnetic tape cassettes. One of the cassettes displays the data-collection form on a CRT, the other cassette records the data keyed in by an operator. When keying is completed, the data tape is rewound, a dial-up link between the computer and the terminal is established, and data are loaded directly into the system.

Dictionaries of both new and existing character entries are edited for correctness, and after an accumulation of new character variables is compiled, the new entries are merged with the existing GRASP dictionaries.

Graphics data input. Maps or photographic reductions of maps are the source documents that provide data for the GRASP graphics data file. The device used to digitize these documents is a high resolution rotating microdensitometer (called a "drum scanner"). A map or photo, either opaque or transparent, is directly digitized on the scanner. The scanner measures the intensity of gray from 0 to 3D (256 levels of gray) for cells of 12.5, 25, or 50 microns square (selectable). Optical density is digitally recorded on magnetic tape where features of a given colour or line weight are represented by their respective optical densities; their positions on the source document are recorded as scan interception positions. Once a document is scanned, contiguous cells of the same optical density form a feature. Simply stated, the scanned image is analogous to a home television picture.

The digitized image is replayed directly onto a storage type CRT equipped with a locatable cursor (or light pen). The cursor is positioned within the boundary of any desired pool, and the identification number of the pool (read off the screen) is keyed directly into the computer. At this point the entire map image has been digitized, and an array of cursor positions with corresponding identification numbers has been entered into the computer. The digitized image is then processed so that only line vectors that completely bound the cursor positions are extracted from the digitized map image. These boundary vectors are outlines of the pools that comprise the data bank. Figure 1 illustrates this entire operation, including the computer overlaying that extracts the appropriate boundary-line vector associated with each identification number. Figure 2 contains a direct print of a photo-reduced map of the oil pools in northeastern Colorado, a print of the scanned image of this photo, and a two-step enlargement of the scanned image to the scale of the original map (1:500 000). There is no easily detectable difference between these photos, and the digital step increments in the enlargement to original scale are smoothed by the grain of the film so that the increments are not detectable.

Data Output

At the present time, interactive graphics type CRT's are the primary means of GRASP data output, although any standard alphanumeric terminal will accommodate the tabular output. When the graphics portion of the system is fully operational, the CRT will serve as the primary output terminal, and output will subsequently be directable to any peripheral graphics device.

In order to enhance the graphics output, geographic constants files composed of digitized political boundaries, major drainage, and standard coordinate systems, will be used. Any of these geographic files can be directly overlayed on a retrieval map.

GRASP is presently operational on the following computers: UNIVAC 1108, PDP-1070; batch-only operation has been executed on an IBM 360/65 and an IBM 370/155. It is expected that in the very near future GRASP will be time-share operational on the IBM 370/155.

Data-base management system for environmental geological data in Czechoslovakia

J. Hruska
Geoindustria Praha
U Pruhonu 32
Praha 7
CZECHOSLOVAKIA

In Czechoslovakia a system has been developed for environmental (spatial) data storage and retrieval pertaining to regional planning, natural environment and economic minerals evaluation. Czechoslovakia's intensive economic activities must take account of further restrictions imposed on new building sites and development areas in order to save natural resources, as yet unexhausted, for generations to come. In the framework of a National System I.S.U. (Integrated National System of Territorial Information), the organizations of geological surveys and Geofond have prepared or are preparing a GEOLOGICAL SUBSYSTEM of the I.S.I. databank. This subsystem has now reached a development and experiment of 10 files, connected by a unique computer system, common minimum data items and common file structure. These files, in various stages of implementation, have a content and purpose as follows:

1. Mineral deposits (including industrial minerals)
2. Geological field maps
3. Waste deposits, undermined areas, etc.
4. Nature conservation objects and areas
5. Potable groundwater sources
6. Mineral ground waters and spas-areas
7. Landslides and other geodynamic phenomena
8. Hydrological and water-engineering data
9. Engineering geological maps and site investigation reports
10. Engineering geological borehole-descriptions and tests.

Every year the Czech Geological Office and Geofond – the Czech documentation state service – complies with hundreds of requests for information on many practical problems, summarizing and furnishing highly relevant records to professional workers who make use of them for specific tasks. More and more quantitatively oriented demands such as in table or graphic form, statistical reviews, etc. have been required recently. This has taken into consideration a project of environmental geological compatible files based on existing information sources in the first stage. Establishing elements and standards of a national spatially conceived system will facilitate dissemination, exchange and correlation of data by computer.

Table 1. Comparison of data-base management systems for environmental-geological information (*available 1972 in Czechoslovakia*)

Systems items	A S T I Czechoslovakia	G I P S Y USA	I M S USA	S A F R A S Canada	S I G M I France	U D B West Germany
1. ORIGIN-Author	Fendrych-LPS VUT Brno	University of Oklahoma	IBM Rockwell (1969)	P.G. Sutterlin J. de Plancke University of Western Ontario	Centre d'Informatique Géologique	IBM Boblingen
2. Information "Know how" in Czechoslovakia	M. Fendrych, Brno J. Hruška, Praha	V. Sattran, Praha	V. Thannabauer, Praha	J. Hruška, Praha	V. Sattran, Praha	CUA Bratislava TERPLAN, Praha
3. Estimated number of users (1972)	11	30	18	30	2	3
4. Price of programs 1972	Kčs 30 000.	$1 500.		$1 500.		
5. Programming language	ALGOL-Genius	Assembler	COBOL, PL1	COBOL	COBOL	Assembler, COBOL
6. Structure of data	"words" + characters	?	free format (or descriptors)	free from format	hierarchical (bite-words)	7 bits-word
7. Minimum extent of memory	32 K	128 K	256 K	128 K	64 K	64 K
8. Type of computer used	Datasaab 21,22	IBM 360	IBM 360/370	IBM 360	IBM 360	IBM 360
9. External Memory	MT/MD/	MD/MT/	MD;MT	MD;MT	MD/MT/	MD;MT
10. Input data	Punch Tape (PC 80,90)	PC;MT	PC(PT)	PC;MT	PC	PC(PT)
11. Indentification (identifiers)	no	labels	labels	COBOL-names	no	key
12. Max.characters/ 1 data item	66	999		99	256	20 AN or 8 N
13. Max.data items/1 record	150				no	"unlimited"
14. Max. characters/ 1 record	2 400 (9 900)	32 768			4040	"unlimited"

After we had experimented with two smaller files in 1970, we created a task group for system selection.

First, information was obtained from a report of CODASYL system's commission (1971) "Analysis of properties general data-base systems". This report concerns ten major systems. From these ten systems only two were actually tested or prepared in regard to areal application:

NIPS/FFS
IMS

Next, information was obtained from Hruska and Burk (1971). After excluding non-available and pilot project systems, there remained only four available computer systems.

Finally, program systems developed and applied in Czechoslovakia were added (see Fig. 1).

Building Raw Minerals File: A Case Study

During the past five years GEOINDUSTRIA n.p., in cooperation with Geofond, collected geological, technical and economic information for about 16 000 quarries, gravel and sand pits, either operating or abandoned, and geologically advantageous outcrops for possible exploitation. This information recorded on unique "open-language" format was compiled by many professionals partly from existing documents (maps, technical reports, mineral reserve calculations, etc.) and partly by field observation.

File content was developed after competent discussions. A standard record form contains 120 data items divided into eight parts (Figure 1).

Mineral Building Raw Materials File

A system design was tested by using programs of the ASTI system. The experiment on 298 records of carbonate industrial deposits of Bohemia and Moravia was very efficient for storage and user-oriented retrieval. Prospective aims of compatibility and integration of files with the official I.S.U. databank has been forecast, so that the input form is now available for both systems, as well as a data base, manuals and formats.

Computerized Engineering Geological Maps

Engineering-geological maps at a scale of 1:25 000 are now being constructed in Czechoslovakian areas designed for both industrial and building development in years to come. These areas are usually covered by a dense network of previous geological records, with about 10-12 reference points per sq. km. One map sheet contains about 1000 boreholes.

Primary records - location and geologic description of the boreholes and exposures, as well as results of laboratory tests of physical properties of grounds - were formalized, punched and processed through the I.S.U. standardized routine. The Engineering-geological Map selected routine was used to select and calculate constructional values - boundaries and thickness of individual beds above sea level and groundwater level.

The program was also helpful in transforming the coordinates and processing both the selected and calculated values so as to assume the shape required for an interpolating routine input. The interpolation routine was that of SYMAP, version V, with graphical output represented by the maps showing contour lines of the respective constructional elements (IBM 360/40). Seven

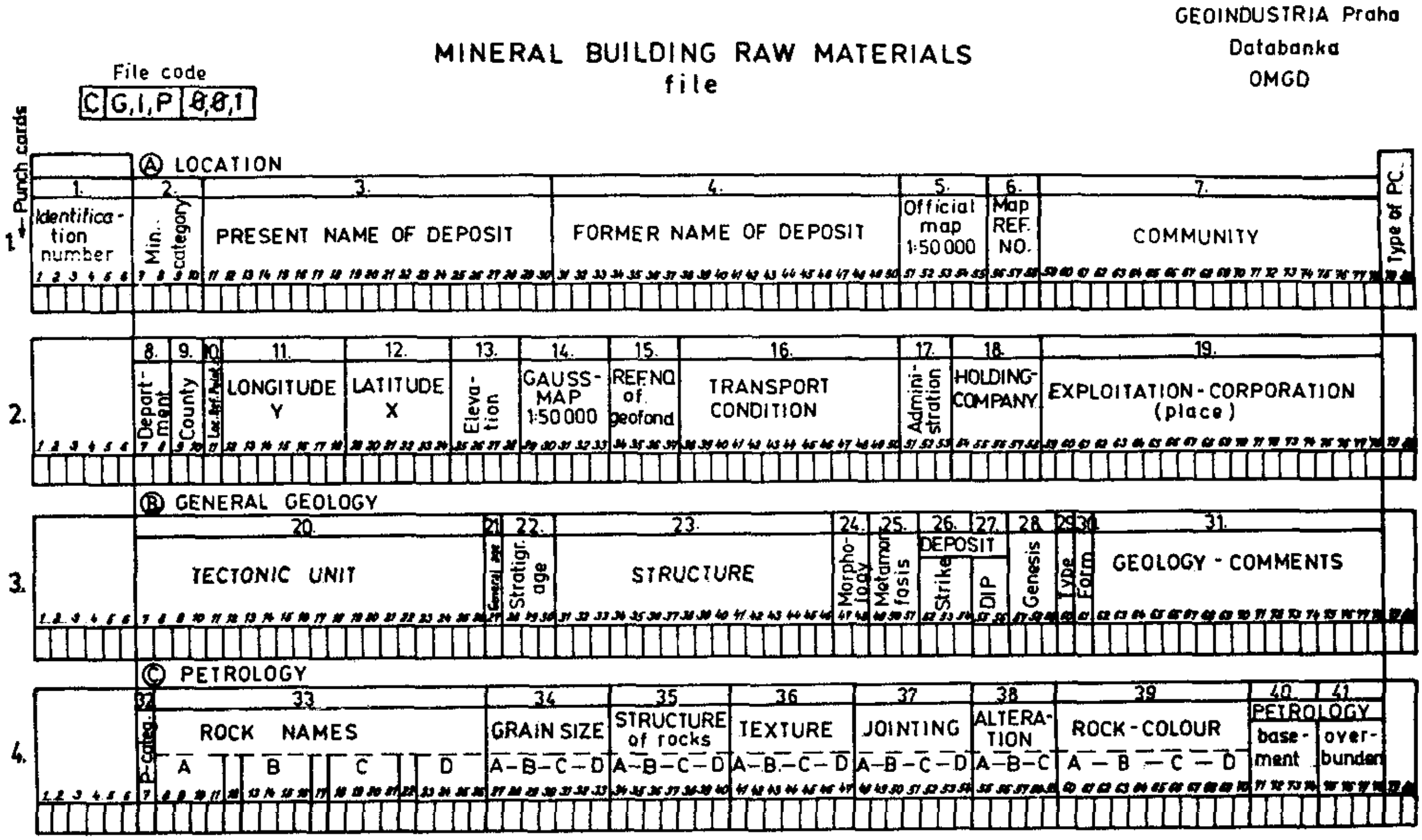


Figure 1. Mineral building raw materials file.

analytical maps of the area were constructed during one cycle (contour lines of pre-Quaternary basement, isograms of Quaternary bed thickness, hydroishyets etc).

Computer-processable data files should be regarded as fulfilling the growing need for easily accessible and readily communicable information, and can be thought of as providing "computerized information potential". The usefulness of such files could be significantly enhanced if the following points are taken into consideration:

1. Content must be meaningful, homogenous and significant.
2. Notation must be internally consistent, meaningful and as efficient as possible for computer.
3. File structure must be as simple as possible.
4. System design must allow sufficient flexibility to permit file restructuring without major systems modification and be adaptable. The output should be tailored to the needs of the user, to statistical subroutines, etc. Databank arrangement is the best goal of such systems.

The problem of notation (coding, man-machine interface) appears to be the first serious problem of the file creator. The meaning of geological data is often more difficult to transmit than the meaning of data from more exact sciences. The primary reason for this seems to lie in the mental habit inherited from geological tradition.

Bibliography

Hruska, J., and Burk, C. F., Jr.
1971 Computer-based storage and retrieval of geoscience information: Bibliography 1946-69; Geol. Survey of Canada, GSC Paper 71-40, 52 p.

Buzkova, H., and Shanelec, V.
1973 Computerization of engineering geological map (in Czech), In: Bulletin of "Hornicka Pribram ve vede 1973" sekce M, p. 465-486.

Fendrych, M., and Fogl, J.
1971 Automated system for information retrieval ASTI-SAAB in Czechoslovakia; Inform. Stor. Retrieval (U.S.A.), vol. 7, 1971, no. 5, p. 245-248.

Hruska, J.
1971 A short review of data processing in the earth sciences in Czechoslovakia; Jour. Inter. Assoc. Math. Geol., v. 3, no. 4, p. 369-373, 2 figs.

Valtr, A., and Shanelec, V.
1971 Regional planning aspects in developing of geological information system (in Czech): in Bulletín Hornicka Pribram ve vede 1971, part M 22, 8 p.

INDEX TO SYSTEMS

Country **Page**

Software System Name or Acronym