

**NATIONAL
ADVISORY
COMMITTEE
ON
RESEARCH
IN THE GEOLOGICAL SCIENCES**



This document was produced
by scanning the original publication.

Ce document est le produit d'une
numérisation par balayage
de la publication originale.

**Interim Report of the Committee
on
STORAGE AND RETRIEVAL OF
GEOLOGICAL DATA IN CANADA**

Compiled by S. C. Robinson

Published by the Geological Survey of Canada as GSC Paper 66-43



G E O L O G I C A L S U R V E Y
O F C A N A D A P A P E R 66-43

S T O R A G E A N D R E T R I E V A L O F G E O L O G I C A L
D A T A I N C A N A D A

Interim Report by the ad hoc Committee of the
National Advisory Committee on Research in
the Geological Sciences

Compiled by
S. C. Robinson

D E P A R T M E N T O F
M I N E S A N D T E C H N I C A L S U R V E Y S
C A N A D A

© Crown Copyrights reserved

Available by mail from the Queen's Printer, Ottawa,
from the Geological Survey of Canada,
601 Booth St., Ottawa
and at the following Canadian Government bookshops:

OTTAWA

Daly Building, Corner Mackenzie and Rideau

TORONTO

Mackenzie Building, 36 Adelaide St. East

MONTREAL

Aeterna-Vie Building, 1182 St. Catherine St. West

WINNIPEG

Mall Center Bldg., 499 Portage Avenue

VANCOUVER

657 Granville Avenue

or through your bookseller

A deposit copy of this publication is also available
for reference in public libraries across Canada

Price 50 cents Cat. No. M44-66-43

Price subject to change without notice

ROGER DUHAMEL, F.R.S.C.

Queen's Printer and Controller of Stationery

Ottawa, Canada

1966

CONTENTS

Page

Preface	
Abstract	

CHAPTER 1

Introduction	
Events leading to formation of this Committee	1
Appointment of Committee and terms of reference	2
Definitions	3
Meetings	4
Appointment of consultants	5

CHAPTER 2

Data Processing in Geology	6
Introduction	6
Application of data processing to geology	6
Difficulties in applying data processing to geology	7
Principles in applying data processing to geology	8

CHAPTER 3

The Storage and Retrieval of Geological Data in Canada	9
Existing status	9
Requirements of a National System for storage of geological data	11
The concept of a Universal National System	13
Factors common to all files of earth sciences	13
A System of linked files	13
Conversion from project files to archival files	14
Selection of data for archival storage	14
Type of processing system to be used	15
Procedures for recording data	16
A National Index to Geological Data	17

CHAPTER 4

Geographic Coordinates	17
Desirable characteristics in a system for defining location	17
Methods of geographic location in use in Canada	18
Geographic coordinates recommended for use in the National System	19

CONTENTS (cont.)

Page

CHAPTER 5

Reference Numbering	21
Reasons for a reference number	21
Current usage in Canada	21
Reference numbering recommended for the National System	23

CHAPTER 6

Digitizing Descriptive Data	24
Introduction.....	24
Factors in recording descriptive data	25
Descriptive data recorded for projects	26
Consistency in recording descriptive data	27
Recommended format for sample descriptions	29

CHAPTER 7

Coding of Geological Names and Terms (C.F. Burk, Jr.)	29
Derived and assigned coding methods	30
Mnemonic codes	31
Recommended coding procedure	31
Conclusions	33

CHAPTER 8

National Index to Geological Data (F. Raynes)	34
Requirements of a National Index	34
Administration of a National Index	35
An existing information system applicable to a National Index ..	36
Recommendations and conclusions	39

CHAPTER 9

Data on Mineral Deposits	40
Introduction.....	40
Uses to be made of the file	41
Existing files	41
Definition of a mineral deposit	43
Development of a format for the National System	44
Level of observation	44
Progress of subcommittee	45

CONTENTS (cont.)

Page

CHAPTER 10

Data on Fossil Fuels and Related Economic Sedimentary	
Deposits (C.F. Burk, Jr.)	46
Introduction	46
Purpose of proposed file	46
Unit of description	47
General approach	48
Nature of data	48
Data entries and general specifications	49
Future work on file development	50

Appendices

I	Members of the subcommittees	51
II	Advantages and disadvantages for data processing of the various coordinates used for geographic location in Canada	58
III	Alberta Oil and Gas Conservation Board unique well identifier	63
IV	Recommended reference numbering format	67
V	Report on the frequency and degree of commonness of the information recorded in field notebooks	68
VI	Sample description code as used by Geological Survey of Canada	70
VII	Coding procedures in National Index to Geological Data	79

Tables

1.	Sample description format by K. R. Dawson	83
2.	Sample description format by N. M. Ediger	84
3.	Mineral deposit record	85
4.	Recommended entries and general specifications for file on fossil fuels and related economic sedimentary deposits	91
5.	A suggested numeric hierarchical code for geologic time units	95

Figures

1.	Sample description format of N. M. Ediger as a spread sheet	96
2.	Indexing document	97
3.	Format of index	98

PREFACE

Minute 20 of the meeting of the National Advisory Committee on Research in the Geological Sciences (NACRGS), April 1965, which established the ad hoc Committee on Storage and Retrieval of Geological Data in Canada, required that a report be prepared for presentation to NACRGS annual meeting in April 1966. It became clear that the ad hoc Committee could not complete its deliberations in one year and in consequence, a synopsis of its principal conclusions and recommendations was presented to NACRGS on April 19, 1966. That synopsis forms the basis for, and is printed as the Abstract of, this Report.

However, in agreeing that a preliminary report be prepared for publication, NACRGS recommended that such a report include background material in addition to the substance of the minutes of the ad hoc Committee and its subcommittees, so that it could be used as a reference by those contemplating establishment of geological data files. For the same reason it has been decided to repeat material as necessary under different headings in this Report.

The following report is thus a compilation of information from several sources, and records the decisions taken by the Committee and subcommittees. Although the Abstract has the approval of the Committee, time has not permitted review of the full report by the Committee except where minutes of meetings are quoted. This Report should therefore be regarded as an interim compilation subject to revision in the final report of the Committee which is due in April 1967.

Acknowledgment is made to Dr. C.F. Burk Jr. who prepared Chapters 7 and 10, to Mr. F. Raynes who prepared Chapter 8, and to Mrs. K.L. Edmond, staff geologist, who recorded the deliberations of the committee and assisted in the preparation of the report. As chairman it is my pleasure to thank all the members of the Committee and the various subcommittees for their time and energy; without this cooperation my task as compiler would have been virtually impossible.

OTTAWA, August 1, 1966.

S.C. Robinson,
Chairman.

ABSTRACT

There is a widespread and growing awareness of the need for machine-processable geological data because of the accelerating rate at which data are being accumulated. Until now however, there has been no uniformity in format or data content in the file systems which have been developed by a number of individual organizations throughout the country for their own particular needs. An increasing number of other organizations are presently considering or planning machine-processable data storage facilities and would welcome guidance or advice in so doing.

For this reason, this ad hoc Committee was set up by N. A. C. R. G. S. and asked to report its recommendations by April 1966. As a result of its work, the committee feels strongly that a National System (i. e. a combination of formats, data content, etc.), which could accommodate files of any organization and facilitate the exchange of data throughout the country is most desirable and if set up would be used by many organizations. Therefore, it recommends unanimously that such a system be established. The committee believes further that the problem is urgent and recommends that every effort should be made to expedite the establishment of such a National System. It also recommends that this report, if approved be published at the earliest possible date and given the widest possible distribution, to acquaint all geologists with the committee's purpose and to enlist the active support of all interested persons and organizations.

In more detail, the committee's findings and recommendations may be summarized as follows:

1. The committee reviewed the advantages and problems involved in developing a single system for storage of the data of the various disciplines that are represented by the National Advisory Committee on Research in the Geological Sciences. Because these data are so diverse and complex and because many potential users are concerned with only a limited field, it was decided that a single system would be too ponderous and complex to be useful. Instead, it is recommended that systems be developed for each of the main files needed, and that these systems be linked by standardization of factors common to all of them. These factors are:

- a. Geographic location
- b. Reference numbers
- c. Method for recording descriptive data of geology
- d. Coding of standard names and terms

Geographic Location The various geographic coordinates and their characteristics are reviewed. The committee recommends that a National System record geographic location in degrees of latitude and longitude to 5 decimal places and suggests a method for recording size of rectangular area and precision of location.

Reference Numbers The report reviews uses of reference numbers and the committee recommends a flexible numbering system that permits incorporation of existing procedures used by various organizations.

Method for Recording Descriptive Data In most instances numerical data in geology are of use only when related to relevant descriptive data and both should be stored and processed statistically. It is essential, therefore, that consistency in terminology and selection of descriptive data be equated to expected consistency of numerical data. It was decided that mineral names as defined in Dana, and fossil descriptions comprising phylum, class and order may be accepted as data. Lack of consistency in the use of rock names makes them acceptable only as conclusions. The report of subcommittee No. 2 recommends that data characterizing rocks be recorded, in addition to the observer's conclusion as to the proper rock name. This procedure was accepted by the committee as a policy that should be applied, where possible, to other essential descriptive terms.

The low level of sophistication implicit in this proposal is considered preferable to degrading the objectivity of numerical data by introducing conclusions in the descriptive data.

The committee recommends that the data required by spread sheets should be recorded in addition to, and in no sense replacing, the normal integrated conclusions and interpretations that are usually recorded by field geologists. The minimal data required by the basic spread sheets may be readily extended by more detailed records required for specific purposes.

Coding of Standard Names and Terms The committee recommends that coding be kept to a minimum and that codes be self-evident wherever possible. A specific, machine-processable, mnemonic code is recommended.

2. It was recommended that a series of systems, linked by the common factors listed in section 1 above, should be designed for the various component disciplines in Geology as the need arises. Files for which the need exists today include:

- a. Chemical analyses of rocks and minerals
- b. Data of mineral occurrences
- c. Data of fossil fuel deposits
- d. Geochronological data

- e. Aeromagnetic data
- f. Gravity data
- g. Isotopic data
- h. Data of hydrogeology
- i. Data of areal geochemistry
- j. Data processing programs ("Program Share Group")

Work by the petroleum industry and government on oil well data storage is so far advanced that the only contribution this committee could make is the suggestion that where feasible the common factors of section 1 be incorporated into the systems used.

Work on a system for storage and retrieval of chemical analyses of rocks and minerals is being undertaken in the Geological Survey following a format that originated in the U.S. Geological Survey.

Existing spread sheets for recording data of mineral deposits have been considered by Subcommittee No. 5 and preliminary consolidation of a sheet for recording field observations and analytical data is well advanced.

The content and general specifications recommended for a file on geologic data pertaining to deposits of fossil fuels and related sedimentary deposits was established by Subcommittee No. 6.

3. It was decided that systems that would be sufficiently flexible to handle the complex and largely descriptive data of geological disciplines could be accommodated on unit record equipment and that these systems should therefore be designed for use with modern computers only.

4. National Index The committee recommends that a National Index to geological data in Canada be established and that data in all forms, published and unpublished, should be indexed. It further recommends that responsibility for preparation and maintenance of this Index be undertaken by the Geological Survey of Canada.

It is recommended that unpublished data be indexed only when offered by the organization having title to them. Moreover the data remain the property of, and in the possession of, each organization and their inclusion in the Index does not necessarily imply that they will be released to everyone wishing to have access to them.

The committee accepted the offer made by Imperial Oil Limited of certain components of their confidential Information Retrieval System that can be modified for use in the National Index.

The system will take the form of a catalogue with data indexed under a number of relevant concepts.

5. It is recommended that the ad hoc Committee be retained on the same basis for one additional year.

- a. to assist and advise in establishment of the National Index;
- b. to carry forward the design of spread sheets for those files for which there is current demand in Canada, and to seek the advice of such consultants as may be necessary to implement related programs;
- c. to act as the clearing house for Canadian representation on international committees in this field;
- d. to provide representatives of geology on national committees concerned with storage of data in the environmental sciences;
- e. to advise on and encourage the application of computer (quantitative) methods to geological data in Canada, and to promote the training of Canadian geologists in the assembly and utilization of machine-processable files; and
- f. to bring in a final report to the National Advisory Committee in April 1967.

STORAGE AND RETRIEVAL OF GEOLOGICAL DATA IN CANADA

CHAPTER 1

INTRODUCTION

Events leading to formation of this Committee

Electronic data processing (EDP) methods have been assuming a progressively more important role in geological research. Initial large-scale geological application of these techniques has taken place in support of exploration for petroleum in western Canada to cope with rapidly increasing files of oil well data. Stimulated by the rapid generation of numerical data applicable to other fields of Geology, general interest had grown to the point where in 1963 each subcommittee of the National Advisory Committee on Research in the Geological Sciences (NACRGS) was asked to present a report on applications of electronic data processing to the annual meeting in 1964. At the same meeting Dr. K.R. Dawson presented a report on a system being developed in the Geological Survey of Canada for storage and retrieval of data on samples analysed in the Survey's laboratories. From these reports it was evident that a large number of geological organizations were interested or involved in establishment of data storage files and that many of them were looking for information and advice.

Having considered these reports, NACRGS decided that real advantages would accrue to the science if it were possible to develop a comprehensive system for storage and retrieval of geological data that could be used by any organization in Canada. The Geological Survey was asked to expand its study with a view to extending its proposed system to accommodate the requirements of the geological sciences in Canada as a whole. As the first step Dr. S.C. Robinson was instructed to make a preliminary study of the feasibility of establishing such a system and to report in April 1965. This study was made with assistance of an advisory group within the Geological Survey and with professional advice of Professor W.C. Krumbein of Northwestern University. Data were gathered on visits to various geological institutions in Canada, to the Bureau of Geochemical Census of the United States Geological Survey in Denver, and to Gulf Oil Co., Research and Development, in Pittsburgh. These visits were supplemented by a questionnaire sent out to about 50 organizations in Canada. At the request of the Mines Ministers' Conference, which was also considering applications of EDP, completion date of the report was advanced to January 1965. This report was presented to the meeting of NACRGS in April 1965 and about 250 copies have been distributed in Canada and abroad. After considering this report, NACRGS recommended establishment of the present ad hoc Committee on Storage and Retrieval of Geological Data in Canada.

Appointment of Committee and terms of reference

Minute 20 of the meeting of NACRGS in April 1965 is as follows: "The National Advisory Committee endorsed in principle the findings in Dr. S. C. Robinson's report on a National System for the storage and retrieval of geological data in Canada, and requested that a committee be appointed to take the necessary steps to develop such a National System in cooperation with interested Canadian organizations, a report on this system to be presented to the National Advisory Committee not later than its next meeting in April, 1966".

With the concurrence of NACRGS, Dr. S. C. Robinson was appointed Chairman of the ad hoc Committee, with instructions to coopt other members of the committee to represent the various disciplines in Geology and to represent organizations in industry, education and government known to be interested and active in data processing. The following scientists consented to serve:

Dr. Robert Assad, Department of Natural Resources, P.Q.
Dr. B.K. Bhattacharyya, Geological Survey of Canada
Dr. W.C. Brisbin, University of Manitoba
Mr. A.E. Buller, Union Carbide Exploration Limited
Dr. C.F. Burk, Jr., Geological Survey of Canada
Mr. N.M. Ediger, The British American Oil Company Limited
Mr. A.S. Gibson, Imperial Oil Enterprises Limited
Dr. S.S. Holland, Department of Mines and Petroleum Resources, B.C.
Mr. R.R. Potter, Mines Branch, Dept. of Lands & Mines, Fredericton, N.B.
Mr. F. Raynes, Cominco, Limited
Mr. D.A. Sharp, Department of Energy and Resources Management, Ont.
Dr. J.R. Smith, Saskatchewan Research Council
Dr. D.R.E. Whitmore, Geological Survey of Canada

Members of the Subcommittees are listed in Appendix I.

Faced with a major task and very limited time, the Committee reviewed terms of reference in Robinson's report which had been endorsed in principle by NACRGS and decided to adopt the following guidelines:

1. The study is restricted to methods of storage and retrieval of Geological data in Canada and does not include retrieval of information from published sources.¹

2. The study is concerned with development of a method or methods of data storage and retrieval that could be adopted by any geological organization in Canada wishing to establish its own files. It is specifically not the purpose of this study to initiate any central system of National Files.

3. Disciplines to be included in the study are those covered by the National Advisory Committee: Mineral Deposits, Fossil Fuels, Mineralogy, Geochemistry, Petrology, Stratigraphy, Palaeontology, Structural Geology, Pleistocene Geology and Physical Methods Applied to Geological Problems.

Definitions

In deliberations of the ad hoc Committee a need for definitions of the four terms, 'system', 'index', 'data', and 'information' was recognized. As used in this report these terms are defined as follows:

'System' - the aggregation of formats, general contents, programmes and other factors needed for a machine-operated file for storage and retrieval of data. Thus a single system for geology would have to provide for storage and retrieval of data of all disciplines comprising the broad field of Earth Sciences. A multiple system might provide systems for each discipline linked by certain components common to all. Specifically the term National System as used in this report refers only to a combination of formats, contents, papers, etc. that would enable all organizations in Canada to store and retrieve data in their own files by a system common to all. In no sense does it imply establishment of a central or national file of data.

'Index' - a catalogue of geological data available in Canada, which refers to the documents in which they are stored and to the organization having possession of them. A National Index would have to be compiled and maintained by a central agency but data would be catalogued in it at the discretion of the organization holding title to them.

¹Information retrieval is international in scope and is currently being developed for English speaking geologists by the United States Geological Survey and by the Geological Society of America who are preparing respectively the Bibliography of North American Geology and the Bibliography of other parts of the World. For French-speaking geologists the Bureau de Recherches Géologiques et Minières in Paris have prepared a series of bibliographic reference cards on World geology since 1959. Other sources of information are library catalogues, accession lists and cards sent out announcing new publications of the Geological Survey.

'Data' - reproducible measurements and descriptions of geological and related phenomena. Descriptive geological observations are acceptable as data only when consistency in use by at least 90% of graduate geologists can be reasonably expected.

'Information' - processed data, published or unpublished, including compilations, conclusions, interpretations, statistics, etc.

Meetings

The Committee met three times. The first meeting was held in Ottawa, July 9, 1965, when the following subcommittees were appointed:

1. On geographic location and reference numbering
Chairman: D.A. Sharp
2. On digitizing descriptive data
Chairmen: S.C. Robinson, D.R.E. Whitmore
3. On mnemonic coding
Chairman: C.F. Burk, Jr.
4. On a National Index to geological data
Chairman: F. Raynes

The second meeting was in Calgary, October 27 and 28, 1965, at which reports of the four subcommittees were discussed, and the following two subcommittees were appointed:

5. On data of mineral deposits
Chairman: D.R.E. Whitmore
6. On data of deposits of fossil fuels and related economic sedimentary deposits
Chairman: C.F. Burk, Jr.

At this meeting the proposed National Index was discussed in principle, and at the suggestion of Mr. A.S. Gibson the Committee was privileged to hear comments by Mr. J.W. Cherry, Chief, Technical Service Information Department, Imperial Oil Enterprises Limited. As an outcome of these discussions, use of the "CAT" System of Imperial Oil, modified as needed for our requirements, was offered and gratefully accepted.

The third meeting was in Toronto, March 9 and 10, 1966, when reports of all subcommittees were reviewed and decisions taken concerning the contents of a final report to NACGRS. One recommendation provided for continuation of the work of the ad hoc Committee for an additional year.

Subcommittee meetings were held as follows:

- No. 1 In Toronto, August 19, 1965
In Calgary, September 17, 1965
- No. 2 In Ottawa, August 5, 1965
In Calgary, October 26, 1965
- No. 3 In Calgary, September 1, 1965
- No. 4 In Montreal, September 17, 1965
In Toronto, November 30, 1965
- No. 5 In Winnipeg, December 16 and 17, 1965
- No. 6 In Winnipeg, December 16 and 17, 1965
In Calgary, January 26, 1966

Contents of the minutes of all these meetings supplemented by a large volume of correspondence provide the basis for this report.

Mention should be made also that at the request of the Annual Meeting of Mines Ministers, S.C. Robinson reported progress of the ad hoc Committee to that meeting in Victoria, B.C., on September 13, 1965, and discussed common problems with a working committee of that organization.

Appointment of consultants

The Committee decided that it was advisable that its recommendations be referred to consultants with experience in data processing in order to ensure that they were compatible with machine processing procedures. The committee was fortunate in obtaining the services of Mr. B.A. Hodson, Director of the Computing Centre at the University of Manitoba. Mr. Hodson attended meetings of subcommittees Nos. 5 and 6 in Winnipeg and of the main Committee in Toronto.

The Committee is particularly appreciative of the tutelage and advice contributed by Mr. J.W. Cherry of Imperial Oil Enterprises Limited during its meeting in Calgary.

As recommended by NACGRS, costs of meetings of the Committee and its subcommittees were borne by the Geological Survey of Canada.

CHAPTER 2

DATA PROCESSING IN GEOLOGY

Introduction

Studies of the earth employ facets of virtually all the physical sciences and for this reason most new instruments and methods developed in those sciences find a use in some branch of geology. Some of these instruments such as the mass spectrometer and the electron probe, used respectively in age determination and study of elemental distribution in ores, have made possible major advances in geological research. In this context, application of data processing methods to storage and retrieval of geological data may appear to have pragmatic rather than fundamental significance. However, our review of their potential for application in Canada suggests that they should provide a major stimulus to geological research and should also make an important contribution to exploration for deposits of minerals and fossil fuels.

As part of the work done by the Committee, a review of the advantages and problems in applying Data Processing (DP) to geology was undertaken. This chapter presents a summary of these advantages, problems and principles.

Application of data processing to geology

Some of the advantages of using of DP were enumerated in the interim report that was endorsed by NACRGS and others were discussed in meetings of the committee. Some of the more important are:

1. Exchange of basic data (as distinguished from processed data and conclusions that are expressed in maps and reports) will permit geology to evolve from a science of isolated investigations, each dependent on limited data, to a universal science like physics, chemistry etc. whose hypotheses, theories and laws are based on data generally available to everyone. As part of this evolution, hypotheses known to be locally true can be tested for validity against data representing a wider range in area, time etc.
2. Availability of a large body of facts consistently recorded should speed development of geological ideas. At present collection of data occupies most of the time devoted to geological research. Release from this burden should confer upon geologists freedom to devote a majority of research time to the exercise of imagination and initiative in the review and development of scientific concepts.

3. Storage of data in machine-processable form, will permit geologists to develop and test theoretical models designed to explain geological processes. Conversely, hypotheses that are the products of qualitative observations in field and laboratory may be tested statistically.

4. For the pragmatic geologist, the widest use of data storage will be selective retrieval of data either in raw form or statistically treated. Moreover, it will be possible to have these data presented automatically in tables, graphs, maps, etc. designed to serve the user's need.

5. A major advantage is the opportunity to compare data of the various disciplines and to compare data from different areas, deposits etc. Such comparisons lead to recognition of trends, associations and groupings whose statistical validity can be determined.

These associations are of particular value to economic geologists for use in assessing potential mineral areas and mineral deposits. In general such criteria have been developed in the past on the basis of an individual's experience; data processing would permit such criteria to be checked against a wider range of data and would also permit the development of others based on frequency of association of various phenomena recorded in the data files.

6. Data stored in machine-processable forms are easily searched and therefore can be used as a catalogue by geologists looking for particular associations. For example it is easy to search oil well data files and find out what wells penetrate a given formation.

7. The explosive increase in analytical and other numerical data, coupled with increase in costs of publication makes it impractical to publish these data routinely as part of scientific papers. By recording these data in a standard format on magnetic tape, however, they may be referred to in the published text and retrieved as a printout or in other form by any reader who has need of them. In addition to substituting for publication, this procedure has the advantage that the reader can obtain the data he requires in the form he prefers without the labour of extracting them from a published text.

Difficulties in applying data processing to geology

Data processing was developed initially for mathematical computation and its evolution has been primarily concerned with handling numerical data. It was only when data processing techniques were extended to handling descriptive data and when in the same period a large volume of numerical data began to be generated in geology, that it became feasible to apply machine-processable techniques to certain fields of geology. Major problems still exist because numerical data in geology are useful to geologists only when complemented by relevant descriptive data. For example a chemical analysis of a rock is useful only if the rock is described

and its geological relationships known. Data processing is particularly effective in the rapid solution of complex mathematical problems and in handling a large volume of repetitive computations. The logic involved in complex geological problems can rarely be expressed by mathematical formulae and few geological problems have enough factors in common to generate repetitive use of data programs.

Other features of geology are difficult to reconcile with the rigid logic of a machine-operated technique. Firstly, geology lacks a definite systematic code of natural laws and well-established theories. Its system is still based on the multiple working hypotheses of T.C. Chamberlin. Secondly, in much of geology classification of descriptive terms has a traditional rather than a logical basis. For example, two of the commonest sedimentary rocks are sandstone and limestone; one is classified by texture, the other by chemical composition. Thirdly, there is a real lack of consistency among geologists, both in their use of descriptive geological terms and in the data they record. Fourthly, because geology makes use of all the basic sciences - mathematics, physics, chemistry, and biology - its data are probably more diverse than those of any other science. Finally, in geology many descriptive terms have genetic or other connotations whose use involves subjective conclusions. The simple case of using such terms as granite, fault, intrusion etc. are examples of this problem. As Agterberg has pointed out "... computing should be based on 'natural populations' only, since computations based on rocks classified by some artificial scheme may easily lead to results which solely reflect an a priori supposition. The use of certain traditional terms in cataloguing data will be unavoidable, but the user of the file should realize that these terms are a necessary evil....".

Some of these problems can be overcome by the greatly increased flexibility of modern computers. However geologists will have to accept some degree of regimentation, primarily to ensure consistency in the data that are recorded and secondarily in the form and terminology used to store them.

Principles in applying data processing to geology

In the course of committee and subcommittee discussions, certain principles were reviewed and those applicable to this study were adopted. Although these are not new nor unique to geology, they are recorded together for the convenience of those who are concerned with establishing storage systems.

1. Files should be restricted to storage of data: interpretations, conclusions, summaries etc. should be excluded. Inclusion of any information that is based on subjective decisions of the observer detracts from objectivity and debases a data file.

2. Descriptive data should be entered at a level of classification low enough to ensure that consistency in use approaches that of numerical data.
3. The system should be designed to facilitate extraction of data in the form most useful to the user; convenience of input or efficiency of machine operation should be subordinate. If the file does not suit the user it will not be used.
4. Geology is only one of the environmental sciences: storage of its data should have as much in common as is possible with methods used in agriculture, forestry, meteorology, oceanography etc.
5. Formats should be designed so that a scientist's time devoted to encoding and decoding is minimized.
6. Any system of data storage in a complex science must be highly flexible and capable of extension to include additional data.
7. Data should be entered into machine storage as directly as possible in order to reduce human errors in transcription etc. Preferably observations should be recorded directly on the spreadsheet used by the key punch operator.
8. There is no technological limitation to file size or file format and these should not be considered as limitations by geologists in preliminary assessment of the basic requirements of a storage system.
9. A minimal requirement of any system for storing environmental data is that it should be capable of presenting the data in graphic forms such as maps, sections etc.
10. Experience shows that a good system can only be worked out by trial and error. It is advisable therefore to try out each component before incorporating it as an integral part of the system.

CHAPTER 3

THE STORAGE AND RETRIEVAL OF GEOLOGICAL DATA IN CANADA

Existing status

Developments in this field have taken place so quickly in the past five years that attempts to compile data on use of machine processable files in geology tend to be outdated before they can be published. This summary is based on replies to a questionnaire sent out in the Autumn of 1964,

supplemented by information supplied by members of the ad hoc Committee. For these reasons the summary is brief and qualitative.

1. Geological data are assembled and stored for two main purposes:
 - a. for purposes of individual and specific projects (project storage);
 - b. for general reference and multiple use (archival storage).

Project storage is widely used by scientists in most geological organizations and includes only those data that the geologist considers germane to his project. Machine processable project storage greatly enhances a geologist's ability to sort, classify, compare and integrate large volumes of data. Because it is tailored for use in a specific project, input and collection costs are minimized. Formats, codes, computer programs, and in large measure terminology, are all unique to the program. As use of project storage increases in each field of earth science, there will be more conformity both in data stored and in format. Both trends will facilitate transfer of project data to archival storage.

Archival storage is less efficient in that it must provide data for a wider range of reference and/or processing purposes. Moreover it must be restricted to data recorded at a level of observation and in terminology such that consistency in use by a large majority of geologists can be assured. These restrictions sharply limit the scope of data that can be put into archival storage. Nevertheless archival storage of geological data is becoming the principal form used in industry and to an increasing extent in government.

This committee is principally concerned with archival storage. It is highly desirable however, that there be as many factors in common as possible between project and archival storage in order to facilitate conversion of data from one to the other. For this reason data stored for various projects and the formats are being studied by the committee.

2. Principal geological files, archival and project, in which machine-processing methods are known to have been applied, are listed below in approximate order of abundance:

- (a) Data from oil and gas wells (several categories)
- (b) Chemical analyses of rocks and minerals
- (c) Areal geochemistry
- (d) Hydrology - wells and analytical data
- (e) Seismic data
- (f) Geochronological data
- (g) Airborne magnetometry surveys
- (h) Areal geology
- (i) Data of mineral deposits
- (j) Gravity data
- (k) X-ray data on minerals

- (l) Isotopic analyses
- (m) Data of glaciology
- (n) Marine geology

This list indicates that machine-processable files are being used mainly where selective retrieval and comparison of large volumes of data is practicable or where extensive computation of data is necessary. Other geological files in which machine-processable methods are in use or are being considered include: palaeomagnetism, rock densities and other physical properties, palaeontology, catalogues of rock and mineral collections, biogeochemistry, sedimentation data, electromagnetic and induced polarization data, coal petrology, diamond drill logs, rock fabric data, stratigraphy, and catalogues of data processing programs. The Committee recognizes the inadequacy of this record and will make a more complete survey in 1966-67.

Requirements of a National System for storage of geological data

In reviewing requirements of a National System there will obviously be reference to some of the principles listed in Chapter 2. To facilitate reference, some of these principles, modified as appropriate, are restated here.

1. The principal requirement is to ensure that all essential data are recorded; the choice of format is secondary. In a few years formats of even descriptive data may be cheaply converted by computer, but missing data can rarely be observed at a later date. Conversely this means that where requisite data are not recorded, they were looked for and not found. In many files provision will be made for direct statement to this effect.

2. The data must be recorded at a level of observation that will assure at least 90% consistency in use. Geologists will use the data of others only if they are satisfied by the consistency, validity and objectivity of observations and/or measurements, analyses, etc.

3. The system should be initially restricted to data and should ensure that there is no possibility of confusion between facts and processed facts (i.e. information). All statements that call for conclusions by the observer should be labelled 'information' and recorded as such. Further, as Dr. C.F. Burk, Jr., pointed out (personal communication, 1966), the distinction between what may be accepted as data and what must be considered information is in part a function of the level of specialization for which observations are made. In limited fields and particularly for a limited group of geologists much greater latitude in what is recorded as data may be permitted than for broader archival use envisaged in use of a National System.

4. Geology is notorious for the profusion and rapid growth of its terminology, a fact that is largely due to incorporation of genetic or other conclusions in its descriptive terms. In records of data to be used as part of a National System, these terms by definition should not be used because they involve conclusions.

5. A National System should provide a checklist of desirable data in each specialized field. This useful function can only be realized by wide exchange of ideas. Moreover the system must provide for changes in the list as experience and evolving geological thought dictate.

6. The system must be designed so that it can be used without reference to a central National authority.

7. The system must be flexible enough so that it can serve effectively the various specialized files of the diverse disciplines that together comprise geology, without becoming so ponderous that it fails through sheer bulk.

8. The system should be compatible with systems used in other environmental sciences such as agriculture, biology, forestry, geography, meteorology, oceanography etc. so that interrelationships of the data of each may eventually be studied.

9. The system should be designed for use of geologists. Where preferences of systems analysts and geologists cannot be reconciled, the geologists must prevail for the simple reason that they will not use the system if they do not like it. A corollary to this is that the time devoted to input by geologists must be kept to a minimum. If the system is to attain its major objectives however, geologists must be prepared to record minimal data required by the system even though those data are not required for their own projects.

10. It must be universal in design so that the system, or at least its major components, will be useful to different types of organizations - mining and oil companies, universities, research organizations and various units of provincial and federal governments.

11. The system must be designed to facilitate its use for the various applications outlined in Chapter 2, and must be flexible enough to provide for uses that may be developed in the future.

12. Any national system must take cognizance of existing systems and attempt to attain all possible compatibility with them. In many cases this may involve minimizing the cost of programming conversion from the national to the existing system and vice versa.

The concept of a Universal National System

A national system for storage and retrieval of geological data primarily defines data that it is desirable be recorded in each geological discipline. Secondly it establishes the level of observation at which consistency in observation can be obtained. If these objectives are met, the data collected by geologists for individual projects become useful to other geologists for quite different projects, and may be used for objective development and testing of geological criteria and concepts.

To facilitate exchange of data it is necessary first to have a national index or catalogue which will indicate what data exist and from which organization they may be obtained. Secondly it is desirable that data be stored in a standard national format or at least in a format that is compatible with the national format. Adoption of a standard format would also permit establishment of standard programs for sorting integration, plotting, comparing, and making statistical analyses of data that could be used by all organizations electing to use the format.

Factors common to all files of earth sciences

The four factors that must form part of virtually all files on earth sciences are: (a) a uniform method of numbering that will provide a unique reference to each observation or sample; (b) a uniform method of recording geographic location; (c) uniform terminology to describe rocks, structures, textures etc.; (d) coding of standard names and terms.

If data in various subject files such as petrochemistry, mineral deposits, fossil fuels, geophysics etc. are to be compared, it is advisable that they should have at least these factors in common. Chapter 4 is devoted to factor (a), and Chapter 5 to factor (b); Chapter 6 records limited recommendations on factor (c), and Chapter 7 deals with factor (d). In addition to use in archival storage it is hoped that these factors may find use in project storage as well.

A System of linked files

The original terms of reference for the ad hoc Committee required it to consider one comprehensive system to accommodate data of all the earth sciences. A review of the problems involved suggested that any such system would be too complex and bulky to be practical. As noted above, however, there are four main factors that all files are expected to have in common, and it was decided to develop a system of individual files for each field of Earth Sciences, linked by these four factors. Data to be stored in each of these files and formats to be used will be decided as the need develops. Four fields in which need is already demonstrated were chosen by

the Committee for development of individual files; these are petrochemistry, mineral deposits, deposits of fossil fuels and certain special aspects involved in geophysics.

Conversion from project files to archival files

Most data to be stored in archival files will come from files established for specific projects. These project files of course will be designed for maximum efficiency in recording and retrieval of data needed for each individual study. It will be necessary therefore to employ machine-operated conversion programs to select the data needed for archival storage and to rearrange them in archival format.

Such conversions are most readily made when the data in project files are recorded in the same mathematical units as those used in archival files. At increased cost, the conversion program can be expanded to include recalculation of data into units used in the archival files, (e.g. if project files record geographic location in U. T. M., these can be converted to latitude and longitude).

The major problem is likely to arise however where project files record descriptive data such as rock names, in terms that embody conclusions by the observer. These are likely to be valid only for a specific project, and to have precise meaning only for those geologists working on that project. It is in this particular field that machine-processing is likely to be ineffective in making conversions. The subtle differences of meaning implicit in use of specific geological terms, and particularly those involving genetic concepts, are often misinterpreted by other geologists and are certainly beyond the capacity of a computer. If therefore, conversion by computer is to be employed, there must be a systematic and logical basis for it and this in turn requires that data be recorded in terms reflecting observations rather than conclusions.

The two principal requirements for computer conversion of project data to archival storage are: (i) that data required for archival storage are recorded in the project; (ii) that descriptive terms used in the project have meaning that can be unequivocally understood by geologists who are not associated with the project.

Selection of data for archival storage

Selection of data should be governed principally by expected uses of the data. For some organizations whose use of data is predictable this is less of a problem but for government and university files it is particularly difficult. There are however, a few broad guidelines.

1. It is usually unnecessary to record data whose importance is short-lived or data that will soon be replaced. Perhaps a safe yard stick would be to exclude data whose value is likely to depreciate seriously in a period of about three years. Examples would be preliminary observations in the field or semi-quantitative analyses in the laboratory.

2. In geology the distinction between data and information is often ill-defined but in deciding whether to include a specific item it may well be judged on the basis of: (a) whether 90% of geologists will accept it as being valid for their own research, (b) if an element of processing (e.g. recording the average of five values) is being considered, is this value likely to change appreciably as other values are determined? Generally, of course, it is preferable to record all values and let the user decide whether the average will suit his requirements. In some cases however, where projects generate large numbers of uniform values it may well be reasonable to record only a few representative values, (c) Can the items be recorded in terms which do not imply a conclusion by the observer?

3. In some instances, it may be useful to record results of computations (e.g. age determination) or an observer's conclusion (e.g. rock name) in the file. In such cases the data on which the computation or conclusion is based should also be included.

4. Once decisions have been reached as to the data that should be stored in the National System, it is important that all geologists wishing to collaborate in building up files of data should be informed, so that all required observations are made. This may mean that collaborating geologists have to record data that they feel they do not require for their own project.

Type of processing system to be used

In the early stages of the Committee's discussions a system that could be used both for unit record equipment and for computer-controlled operations was considered. This was to be based on storage of data in punched cards in highly condensed form by use of hierarchical coding. Such systems were in fairly common use, particularly for storage of data of individual projects.

The committee decided that although coding and decoding were not a great burden to the few individuals concerned in a project, the complex codes needed for the diverse descriptive data of geology would impose an unacceptable deterrent to use of a National System. A review of equipment available in Canada showed that unit record equipment is now almost always associated with computers and that modern computers are capable of handling data so rapidly that condensation of data by coding is not essential. Furthermore it became clear that storage of data on magnetic tape or discs is highly desirable because punched cards are bulky and must be stored in a controlled atmosphere.

In view of these factors and of the rapid advances being made in computer hardware and software, the committee decided that no technical limitation of file size or format should be considered in determining desirable contents of the National System. It was decided that for convenience in designing formats, the standard 80-column card input would be used, but that a systems analyst would be employed to design the system for computer operation.

Procedures for recording data

Because transcription of data is a major source of error in developing machine-operated data files, it follows that there should be a minimum number of steps between the first record of data and its entry into punched cards or other machine-operated media. Errors are reduced by use of well-designed formats used to record data and also as source documents for the punch operator. Field tests are now being made with pocket punches and with 'mark sense' formats, both of which may be read by machine and their contents automatically punched into cards. Both methods provide real advantages in handling data but have the disadvantage that errors in first recording the data cannot be checked until a printout is available, if then. It is wise to provide a cross reference from these records to the geologist's notebook. It is perhaps worth mentioning that looselead notebooks whose pages are edge-punched cards are proving their worth in expediting assembly of field information into maps and reports.

A less direct method employs special sheets for recording data on which the observer checks off the appropriate items and/or fills in numerical data. These sheets serve as source documents for the punch operator and are easier to check for errors by the observer. They too are used to supplement rather than to replace the usual written field notes.

In laboratories spread sheets suitable as source documents by the punch operators are widely used as the first records of data. There is also a trend to having instruments punch results directly into punch cards rather than to record them on dials or graphs. In most cases instrumental results must be corrected for various factors to produce the required analyses. This is readily done by computer programs based on data recorded in the punched cards. In a few exceptional cases, a small computer is built in to the instrument so that its output is the corrected analysis.

Procedures for recording data are of major importance to successful development of machine-processable data files in Geology both to ensure accuracy and what may be more important, to reduce effort needed for input to a level at which most geologists will be willing to use them.

A National Index to Geological Data

In early discussions on the form that the National System should take, consideration was given to duplicating 'master cards' and using these as a means of indexing the data stored in the system. The Committee discussed this proposal in some detail and decided that an index was desirable but that the proposed method was restrictive.

It was decided that an index to geological data in Canada should be a catalogue of all published data and any unpublished data that might be made available voluntarily. In general the index would not contain data but would indicate the kind of data available and the source from which they could be obtained.

It was further decided that the National Index should be separate from the National System for storing data but should catalogue data stored in all forms including any stored in the format of the National System. Subcommittee No. 4 under the chairmanship of F. Raynes was established to develop such an index.

CHAPTER 4

GEOGRAPHIC COORDINATES

Desirable characteristics in a system for defining location

Location is the one major factor that is common to virtually all geological data and observations. It is also likely to be the most frequently used reference in retrieval of data from files. As J.A.C. Fortescue (personal communication, 1966) has pointed out, however, many problems require study of data not just from a point source but from a prism extending from the atmosphere down through the biosphere and hydrosphere and into the lithosphere. Location therefore must be the common key to files in widely divergent disciplines. For these reasons it is important that any system for recording geographic location in geology be used in, or be compatible with, systems used in other environmental sciences.

Other factors that are important in the selection of a system for defining location in data processing are that it should:

1. be Canada-wide and preferably World-wide in scope.
2. be capable of defining a point as well as an area.
3. contain a minimum of discontinuities between individual coordinate grids and thus be useful for areal and regional computations.
4. be amenable to use on x-y plotters.

5. be readily usable for measurement of coordinates from points plotted on a map both by use of a simple romer and by use of a digitizer.
6. be available on current maps of Canada.
7. be widely known and understood.
8. facilitate reoccupation of a point where a sample was taken or an observation made.

None of the present methods of recording geographic location in Canada fulfil all these requirements and in fact it is improbable that a method can be designed that will fulfil them all. Furthermore, scientific requirements are complicated by legal requirements of land description that are obligatory in some Provinces for defining the location of oil wells, water wells and other features. These legal surveys are continuous and consistent only through parts of the three Prairie provinces.

Methods of geographic location in use in Canada

The methods currently in use in Canada with the number of desirable characteristics satisfied by each (in parentheses) are as follows:

- | | |
|---|-----|
| 1. Geographic coordinates of latitude and longitude | (5) |
| 2. The National Topographic Series | (3) |
| 3. The Township and Range System | (2) |
| 4. Marsden squares | (1) |
| 5. Universal Transverse Mercator Grid | (5) |
| 6. Vertical air photographs | (2) |
| 7. The ARDA system | (4) |
| 8. Local rectangular grids | (2) |

A further point of importance is the selection of coordinates that are most widely applicable to the various map projections that are in common use in Canada. These projections are:

1. Transverse Mercator projection
2. Polyconic projection
3. Lambert Conformable Conic projection

Of these the Transverse Mercator and the Polyconic projections are commonly used at scales of 1 inch = 8 miles, 1 inch = 4 miles, 1 inch = 1 mile, 1:250,000, 1:50,000 and 1:25,000 and larger scales. The Polyconic projection has some advantages in preparation of maps, but distortion at the edges of map-sheets is greater than in the Transverse Mercator with the result that Transverse Mercator has now been adopted by the Topographical Survey and the Geological Survey of Canada. The Lambert Conformable Conic projection is used for maps representing larger areas of the earth's surface, at scales of 1:1,000,000 and smaller.

On all these projections, meridians are straight lines and parallels of latitude are curved. The Universal Transverse Mercator grid is a rectangular grid used on Transverse Mercator projections but cannot be used on Lambert Conformable Conic projections.

Brief descriptions of the various coordinates with comments on their advantages and disadvantages are contained in Appendix II.

Geographic coordinates recommended for use in the National System

Subcommittee No. 1 under chairmanship of D.A. Sharp was asked to study advantages of various coordinate systems and to recommend one for use in the National Index. The committee was also asked to make recommendations on recording the vertical coordinate relative to sea level; to consider methods of digitizing location of areas as well as points; and to discuss methods of indicating the precision with which a point is located.

After two meetings and extensive correspondence, the following recommendations were made by the subcommittee under cover of a letter from Mr. Sharp dated 19 October 1965:

1. Geographic location should be defined by latitude and longitude in degrees to five decimal places.

With regard to making provision for the recording of U. T. M. coordinates the subcommittee recommends that consideration be given only to latitude and longitude for a National Index. It should be stressed that each organization may use internally whatever legal location system is most suited to its particular needs. Presumably the petroleum companies primarily would use the Dominion Land Survey system and the Geological Survey of Canada would use the Universal Transverse Mercator system or modified version thereof. However, for a National Index, it was agreed that a common denominator should be resolved and that latitude and longitude be adopted as the standard.

2. Location may refer to a point or an area.

In the case of area, it was agreed that the subcommittee was not qualified to recommend a system that could describe irregular shaped bodies. As a result, it is recommended that areas be indicated through the use of significant figures i.e. an area would be indicated by padding with zeros and the last significant figure in the five decimal places would be recorded in a separate column following the latitude and longitude description.

If the location of a point is not accurately known, a similar method of denoting accuracy through the use of padding with zeros and indicating the significant figures would be employed. As such a system could create problems in distinguishing between an area and an inaccurate point location for the latter, a standard procedure such as recording a "1" in the fifth column could be used.

It is further recommended that at the next meeting of the ad hoc Committee, the precision factor of five decimal places which permits a horizontal accuracy of greater than five feet be reviewed.

3. Geographic location should consist of an x-y coordinate.

The third dimension or "z" coordinate is not generally consistent for all disciplines and will vary depending upon the nature of the observation. As a result, the subcommittee agreed that it would be a misnomer to include such in geographic location which, practically speaking, refers to a two dimensional observation. However, at the request of Dr. Bhattacharyya, the problem of "z" coordinates will be referred to the main committee for further discussion.

After discussion of these recommendations at meetings of the main Committee in October 1965 and March 1966 some minor modifications and clarifications were proposed which are reflected in minute 4.1(B) of the meeting of March 9 and 10, 1966, as follows:

"Mr. Sharp summarized recommendations of the subcommittee which were agreed upon at the Calgary meeting, particularly those to the effect that: latitude and longitude be employed for geographic location within a National system; data collected over an area be located by the latitude and longitude of the approximate centre of the area, using significant figures to indicate approximate areal extent.

Mr. Sharp suggested that if the number of significant figures is employed to indicate areal extent, we must be able to discriminate between an area and an inaccurately located point. Mr. Hodson pointed out that since a point is an area of zero dimensions, it would be possible to describe either a point or an area by latitude and longitude, with an extra digit to indicate area in the case of a square.

The committee recommended that the consultant investigate use of the southeast corner for specifying geographic location of areas.

Following further discussion by Dr. Robinson and Mr. Hodson, it was agreed that the area of the square could be indicated by the number of significant figures used to define it -- the smaller the number of significant figures, the larger the square. For example

08930000, 5120000x would indicate a 'square' 1/10 degree on each side whose S.E. corner is 89.3 degrees west and 51.2 degrees north. The precision with which this point is defined would be indicated by the digit in a column 'x' immediately following the 15 columns used for latitude and longitude."

It was agreed that the precise method for recording latitude and longitude of a point or area be referred to a consultant for advice before taking a final decision. A proposal to use the National Topographic System modified by one degree squares, as an alternate system for geographic location by latitude and longitude was also referred to a consultant for study.

CHAPTER 5

REFERENCE NUMBERING

Reasons for a reference number

There is a remarkable disparity of views on the role to be played by a reference number in storage and retrieval of data. Primarily a reference number is a means of uniquely identifying a specific item of data. It is used in all documents as a means of identification and as such may be used to organize the various items of data in a file into compartments suitable for retrieval.

However there is a strong compulsion to make further use of a reference number for such purposes as identifying: (a) the organization that assigned it; (b) the decade and year of assignment; (c) the project for which it was assigned; (d) the individual responsible for the data; (e) the class of data to which it refers and even the geographic location of observed data. To most geologists a sample number is analogous to a reference number and to many, it has much in common with a library call number. Both the library and sample numbers serve to classify units as well as to identify them.

In machine-processed data storage, the reference number is like a tag which always accompanies the item of data and serves to identify it in the computer's memory. It also serves to link items of fact relating to a single observation point or sample.

Current usage in Canada

Major users of reference numbering each have their own system for assigning numbers. These systems range all the way from simply issuing numbers in sequence, to the well identifier system recommended by the Canadian Petroleum Association and based on geographic location of wells by the township and range. The type of numbering used varies with the size and complexity of each agency. If it is feasible to have all numbers issued by a single authority within an agency simple sequential numbers provide

unique references with a minimum number of digits. However, in an agency with a large number of project leaders operating independently in the field, sequential numbering will not work and even the issuance of blocks of sequential numbers by headquarters has proved to be highly unsatisfactory.

As soon as simple sequence is abandoned, reference numbers must reflect data of some kind. Commonly sequential numbering is prefaced by alphabetic or numerical groups that indicate a specific subdivision of the agency, a specific project or project leader and commonly two digits indicate the decade and year in which the number was used. Such prefixes may also indicate the kind of data (e.g. seismic, isotopic etc.) or in fact may include some part of the actual data.

As examples of methods of reference numbers as applied to samples, it is worth considering the two systems in use at the Geological Survey of Canada. One system is used as an interim measure in the field, and a second system is used for permanent record of those samples on which analyses, or other work worth retaining has been done. Because a large percentage of samples that have been collected in the field will be discarded and not recorded permanently, field numbers are not satisfactory in the permanent record.

The GSC field numbering system comprises the following:

Decade and year	2 digits
Collector's alphabetic code	2 letters
Sequential numbers of samples ¹	up to 6 digits

This system has worked well over the years and caters to the needs of individual project leaders.

The system for permanent record is more complex and has to be more rigid. It consists of 11 digits as follows:

Decade and year of collection	2 digits
Code number of project	3 digits
Station number ²	3 digits
Specimen number ²	3 digits

¹ Considerable latitude is permitted here to provide for inclusion of station numbers, of letters to indicate assistants who collected the sample, etc.

² A zero in the first position of the station number indicates that a station number is not being used in this project and that the remaining 2 digits plus the 3 assigned to specimen number are combined to provide 5 digits for specimen number. This system provides for those geologists (e.g. stratigraphers) who wish to record many samples from one station or locality and also for others (e.g. geochemists) who collect samples individually from many points.

This system provides for 999 projects in each year and in each project for 899 stations with up to 999 samples per station, or alternatively if a station number is not used, for 99,999 samples per project per year. Because some geologists are concerned with more than one project in the year, permanent reference by project is preferred to reference by project leader.

Possibly the most complex geological numbering system currently in use in Canada is the one adopted for oil well numbering by the Canadian Petroleum Association in collaboration with the technical committees of the Mines Ministers' Conference. The format is based on the location of wells by the township and range system and as such, is applicable in those areas where that grid system is available. It may also be used for numbering wells located by the British Columbia modification of the National Topographic grid and by the Northwest Territories grid. Details of this system as adopted by the Oil and Gas Conservation Board of Alberta are included as Appendix III.

Reference numbering recommended for the National System

Subcommittee, No. 1, under chairmanship of D. A. Sharp was asked to consider the reference numbering system that would be most suitable for use in the proposed National System. The following report and recommendations of that Subcommittee was adopted by the ad hoc Committee:

1. Reference number, as considered, is applicable to a data system and not to a National Index.
2. Reference number should be unique.
3. Reference number may be meaningful to the contributor but should have no usefulness as data in the system. Where useful data is included in the reference number, this data must be included elsewhere. It is unlikely that reference numbers will be used for sorting or retrieval, as this is most commonly done on geographic location, class, or discipline, all of which are considered data and hence are independent of the reference number.
4. Each contributing organization must maintain a consistent reference number format.

It would be preferable for such a format to coincide with the format or system being used in that particular organization.

5. A prefix or "key" should precede the reference number. This prefix would designate the reference number format and, where required, the name of the contributing company or organization.

Where several companies use the same format but have reference numbers referring to different data, different prefixes would be used. Similarly, if one company uses more than one format, a different prefix for each format would be required.

Since the prefix would be assigned by a central organization, it is conceivable that by predesignating specific columns broad classifications could be established, thus making the key more meaningful; if a 4 column prefix were used, column 1 could show whether the information pertains to geology, geophysics, geochemistry etc.

A cross reference should be provided to indicate the reference number formats and the companies or organizations using that format.

6. Flexibility in reference numbering is desirable. Although a flexible system incorporating existing reference numbering schemes is recommended, a minimum number of formats would facilitate a greater ease of operation in any system.

In summary, as it is difficult to define what data is necessary for any particular discipline, the committee recommends that a reference number be adopted which consists of a prefix, the number of columns to be determined by estimated usage, and a number assigned by the contributing company or organization consistent with its internal requirements. The maximum number of columns or digits for this number should be stipulated after a further review of existing systems.

A reference numbering format recommended by Subcommittee No. 1 is included as Appendix IV.

It is probable that requirements of reference numbering from the viewpoint of a systems analyst will be referred to our consultant.

CHAPTER 6

DIGITIZING DESCRIPTIVE DATA

Introduction

It is only in the past two decades that numerical data in significant volume have been available in geology. A large majority of geological data however are still qualitative and descriptive. Moreover, few numerical data have significance for geologists unless they are accompanied

by descriptions of the relevant samples or geological environment. An isotopic age of a rock for example, is of little use in geology unless it is accompanied by descriptions of the rock, its environment and field relations, its geographical location, the mineral or minerals dated, etc.

Factors in recording descriptive data

As was pointed out in Chapter 2, if descriptions are to be digitized and recorded with numerical data for use in comparative and descriptive studies, the descriptive terms used must comprise a classification system that should approach, as closely as possible, the logic, consistency, and coherence of a numerical system. Examination of descriptive terms in geology indicates that few such terms have a systematic basis or even unique and agreed definitions. Even in those cases where a unique definition exists, there may well be need for each geologist to draw conclusions in order to use the terms.

The following groups of terms most nearly meet the requirement of unique and agreed definition:

Mineral names (as defined in Dana's System of Mineralogy)
Stratigraphic names (as defined by the American Commission on
Stratigraphic Nomenclature)
Fossils (described by phylum, class and order)
Common structural terms (fault, syncline etc.)

For many other groups of geological terms, it would be difficult to cite one authoritative definition which might possibly be accepted by Canadian geologists. It is particularly difficult to find any basis for agreement in the use of rock names, especially for names of igneous and metamorphic rocks. Even when reasonable classifications are developed, experience shows that they are not used. Rock textures, such as lineation schistosity, rodding and others, are also difficult to define uniquely.

A large number of geological terms are based on genetic or other concepts that call for a subjective conclusion by the observer. Such terms cannot be logically used in recording data. If terms used in recording data are to have unique and agreed definitions and are not to call for a subjective conclusion on the part of the observer, few entries in a glossary of geological terms would be acceptable. To illustrate this point, how many of the following terms would qualify: granodiorite, igneous, regolith, lineation, unconformity, reef, pegmatite dyke, metamorphosed, replacement, plutonic? These terms are used with assurance by most geologists to describe in capsule form their integration of observations of many kinds into a conclusion. They are so accustomed to doing this that they rarely record the observed facts on which these conclusions are based, and if asked will usually refer to these conclusions as facts and so record them in their notebooks.

Use of terms in a record of data can be justified only if consistency in use by geologists is high enough that one geologist is willing to use another's description as a basis for drawing his own scientific conclusions.

It is a matter of experience that almost all geologists prefer to make their own observations in the field rather than accept those of another, whereas they are usually content to accept numerical data. The few tests of consistency in use of descriptive terms that have been made indicate that there are wide disparities in the use of terms by different geologists to describe the same phenomenon.

Unfortunately, knowledge of the subtle shades of meaning conveyed by each of the many geological terms has become accepted as something of a hallmark of geological sophistication, and most geologists are unwilling to forgo their use. Clearly, rigid restriction in data files to terms that are uniquely defined and free from subjective conclusions would severely limit the application of data processing procedures to geology.

Three possible solutions to these problems are:

1. To review existing terms, select basic ones, define them precisely and require that for use in data storage they be used as defined. Definitions should be in terms of observable facts so that subjective conclusions are minimized. Flexibility could be given to such terms by modifying them with approved adjectives.
2. To classify terms into a few major groups based on broadly recognized geological concepts and then to amplify them by description within each category. A good example of this is the method developed in the United States Geological Survey Bureau of Geochemical Census and modified for use by the Geological Survey of Canada for the Geodat file, where rocks are first classified as igneous, metamorphic or sedimentary and then separately described by geological form, grain size, texture, mineral content and finally the rock name suggested by the observer is recorded.
3. To determine and record the actual observations that are used by a geologist in choosing a specific descriptive term. These data can be supplemented if necessary by recording the term chosen by the observer. An example of this alternative as applied to sample descriptions is recorded later in this chapter.

Descriptive data recorded for projects

Little is known about observations that are commonly made by geologists in the field. Normally these vary with the geological environment, the type of project and the training and interests of the geologist concerned.

In an attempt to learn some of the facts that are most frequently recorded, a statistical study of notebooks from five Geological Survey of Canada field projects was made by Dr. F.P. Agterberg. His report forms Appendix V.

Where field data are being recorded for a given project involving a team of geologists of limited size, agreement on exact meaning of selected geological terms can be reached. Those terms can then be used consistently in recording data for that project. Two current examples of such field projects in which data are being recorded on spread sheets, are Operation Pioneer being conducted in the western Canadian Shield by Professor W.C. Brisbin of the University of Manitoba and a major study of the Coast Range mountains in British Columbia by Dr. J.A. Roddick and others of the Geological Survey of Canada. It is expected that both these studies will contribute materially to development of modern methods of recording descriptive data in the field. It is doubtful, however, that it will be possible for these data to be transferred to archival storage in a form that will have precise meaning to geologists other than those taking part in the projects.

It is clear that descriptive data in geology will be particularly difficult to record digitally for archival storage and computer use. In these circumstances it seems advisable to concentrate first on those descriptive data that are necessary to complement numerical data. Numerical data measured in the field include a few physical and chemical measurements, structural measurements and some specialized physical measurements made from the air. However, the large majority of numerical data in geology are generated by laboratory studies on samples brought in from the field. Thus it is logical to concentrate first on digitizing geological descriptions of these samples. Accordingly the remainder of this chapter is devoted primarily to such descriptions.

Consistency in recording descriptive data

Most studies that generate numerical data involve systematic analyses or other procedures. For this reason the numerical data produced are usually consistent both as regards the actual data produced and the form and units in which they are recorded. Similar consistency is essential in the descriptions of the samples. Four general requirements must be met if descriptive data of a sample are to be comparable to numerical data. Firstly the most important requirement is to ensure that all the necessary descriptive data are reported with each sample. Secondly it is necessary to use descriptive terminology that will be understood by at least 90 per cent of geologists and will be free from conclusions by the observer. Thirdly a standard format is necessary to facilitate processing the data and incidentally to act as a checklist. The final factor is development of procedures for recording data in the field that will not impose an unacceptable burden on field staff.

It is common practice among geologists to limit descriptions of samples to those data needed for their own projects. If however, numerical data on these samples are to be used for general purposes, it will be necessary to record certain minimal descriptive data. On the other hand, many projects will require records of more than this minimum. A further complication is that the data ascertained in the field are likely to be supplemented by more specific descriptive data (e.g. less common minerals) that can be determined only in the laboratory. The problem is therefore, to decide what descriptive field data are needed for general use, and to make provision for recording of such additional laboratory descriptions as may be useful.

As mentioned earlier in this chapter, geological terminology is largely unsuited to recording data because few terms are uniquely used and most require conclusions by the observer. The possible methods of dealing with this problem were reviewed earlier in this chapter.

It is useful, but not essential, to develop a standard format for recording descriptive data. Such a format facilitates systematic observation, serves as a checklist, and facilitates transfer of data to punched cards. Wherever possible, formats should provide for a mere check-off of one of a small number of alternatives and for recording any related numerical data such as dip and strike. However provision must be made for use of plain language or mnemonic-coded language, to record minerals present or other factors in which choice of alternatives cannot be restricted to one of a small group. If consistency in the first two requirements can be achieved, data from one format can readily be converted to another by computer and hence consistency in the format used is more a matter of convenience and reduced costs, than it is a matter of necessity. In recording additional data for specific projects separate check-off sheets can be used to supplement the basic sheet and their existence can be indicated by a flag or other device on the basic sheet.

Procedures for recording data are being changed progressively as new techniques become available, and the old method of continuous notes in chronological order is giving way to separate records of notes on loose-leaf sheets, on standard forms, on edge-punched cards, on check-off forms, on cards punched in the field, on tape-recorders and in forms that can be converted to punched cards by an optical scanner. For restricted descriptive data that can be readily recorded for use in computers, check-off forms that can be scanned optically or that can be used as source documents for the card punch operator have real advantages in speed and simplicity of use and in minimizing transcription errors. In the present state of development however, it will be necessary to supplement such records with normal notes either written or recorded on a tape recorder. Many field parties consist of a geologist and an assistant and it is usually the assistant's duty to record the check-off data while the geologist records the notes.

It is expected that the major part of long reference numbers will be gang punched and that only the last five or six digits need be recorded in the field. Location of stations, sample points, etc., will normally be plotted on a map and these can be digitized directly in the office by a coordinatograph.

Recommended format for sample descriptions

Subcommittee No. 2 reviewed the general problem of digitizing descriptive data for samples and decided to devise a format on the basis of an 80-column card for recording field data on a sample taken for more intensive investigation in the laboratory. It was recognized that the resulting format could also serve as a basis for field observations and it was slightly altered for this purpose. Two versions of this format one by K. R. Dawson and one by N. M. Ediger are reproduced as Tables 1 and 2. In both cases column numbers are in parentheses whereas the position number in the column is not. The format of Ediger as it would appear as a spread sheet is illustrated in Figure 1.

A brief code using only 9 columns but based on the observer's primary conclusions concerning general class of rock type or other substance, and thereafter using different descriptive data heading dependent on the first classification, has been in use for chemical data files in the Geological Survey of Canada for two years. This code was modified from one that was developed in the Geochemical Census Branch of the U.S.G.S. in Denver. It is reproduced as Appendix VI.

CHAPTER 7

CODING OF GEOLOGICAL NAMES AND TERMS

(C.F. Burk, Jr.)

In any large data file coding procedures are generally considered necessary in order to reduce clerical effort and storage requirements and to streamline the search and retrieval process. During development of punched-card data files, coding was a particularly important aspect because of the 80-column restraint on formats and the necessity of using numerics for efficient sorting and retrieval. Such files were commonly heavily coded, and although the coding greatly facilitated machine processing of data, output tended to be so cryptic that users were discouraged from using the data, thereby defeating the basic purposes of establishing the file in the first place.

Fortunately for all concerned, particularly users, the necessity for heavily coded input/output formats has been largely eliminated with the

development of modern, sophisticated computers. Coding is still an important aspect of data processing within the computer, but it is no longer necessary for the geologist either to record data or to receive output in a heavily coded form. In a user-oriented system (the only type worth having) the proper technique is to record data in a logical form in the geologist's language; have coding, where required, done either by a technician or by the computer; and have the computer decode output where necessary.

Derived and assigned coding methods

Although heavily coded input/output formats are no longer a necessity, the need for some degree of coding is apparent. For example, in printing output tables it is commonly desirable to confine columns to a limited width, in which case words have to be abbreviated (coded) to fit in the allotted space. Two general approaches to coding are possible, one based on derived methods and the other on assigned methods (Bourne, 1963, p. 46, 50).

Derived coding methods generate a code directly from the original word without the necessity of looking up some item in a table or index. Assigned coding methods, on the other hand, require the use of detailed tables for both the coding and decoding operations. Although the latter methods are exceedingly flexible, in that any unique symbol or group of symbols can be conceived to represent something else, they require a central authority to establish the codes, and the distribution of a document to all users in order for them to use the codes. In view of the general nature of the proposals in this report, which involve independent recording and use of data by a wide and variable number of organizations and individuals across the country, it is clear that the first-mentioned, derived coding methods are the more desirable for use in such a situation.

The main defect in derived coding methods is that with any substantial population of words, none of the methods provides a completely unique set of code words (Bourne, 1963, p. 49). Various schemes provide different degrees of uniqueness, depending on the scheme used, the population of words being coded and the word size (the longer the code word, the higher the degree of uniqueness). When two or more words reduce to the same code word, some arbitrary change must be made by a central authority to distinguish the coded words. For example, in the method recommended here, Pliocene and Paleocene each reduce to the 5-letter code "PLOCN". To distinguish these, Paleocene could be arbitrarily coded to, say, "PLECN". Needless to say, if there are a great number of duplicate codes, the value of having a derived method would diminish and the method would approach that of an assigned code. A certain number of duplicates, however, cannot be avoided.

Mnemonic codes

An obviously desirable characteristic of any code used in a public file is that it can be mnemonic (helping, or meant to help, identification of the word being coded). The ad hoc Committee early decided that all codes, where practical, be mnemonic, and several such coding procedures were investigated.

The creation of a mnemonic code generally involves removing the relatively most commonly used letters from the word, thereby leaving the least commonly used, most distinctive letters in the code word. The first letter of the code word is never deleted.

The order in which letters are removed from the original word has an important bearing on the mnemonic quality of the code, for derived coding methods, on the degree of uniqueness possible. The committee is not aware of any studies to indicate the frequency of letter usage in geological names and terms as such. Studies have been made, however, for such word categories as subject words, proper names, and continuous text (Bourne, 1963, Table 3-3). Four such rankings for continuous text cited by Bourne, and two rankings by Webster and Franklin cited by the Permian Basin well data system, collectively provide the basis (with one minor exception) for the rankings used in the coding procedure recommended by this committee.

Recommended coding procedure

The procedure recommended here is a derived, mnemonic, coding method. It has been used and proven in at least three large files, the Permian Basin well data system, the Ontario well data system and the Geological Survey of Canada petrography-geochemistry-mineralogy data file. By following the prescribed rules, any individual or computer can reduce a word to any desired length. The code will resemble the original word to a greater or lesser extent, depending on the length of the code word relative to the original word. All persons and computers will independently arrive at the same code for the same word.

Ranking of letter usage -- By considering the six rankings previously cited, the following composite order is obtained, in decreasing order of usage:

1. E	7. S	13. U	20. B
2. T	8. R	14. M	21. V
3. A	9. H	15. F	22. K
4. O	10. L	16. Y	23. X
5. N	11. D	17. W	24. J
6. I	12. C	18. G	25. Q
		19. P	26. Z

In order to have a somewhat more easily remembered order and to take into account common (though by no means universal) abbreviation practices, the composite order has been modified by placing the vowels, (A, E, I, O, U) at the beginning, followed in order by the letters W, H, Y and double letters. A further modification of the above order must be made with respect to letters 20 to 22 (B, V, K) in order to conform with the established system (because of the low ranking of these letters, this modification would only rarely affect a 5-letter code word). The final ranking for use in the recommended method therefore becomes:

- | | | |
|-----------------------------------|-------|-------|
| 1. A | 10. T | 19. G |
| 2. E | 11. N | 20. P |
| 3. I | 12. S | 21. K |
| 4. O | 13. R | 22. B |
| 5. U | 14. L | 23. V |
| 6. W | 15. D | 24. X |
| 7. H | 16. C | 25. J |
| 8. Y | 17. M | 26. Q |
| 9. Double letters
(delete one) | 18. F | 27. Z |

RULES FOR CODING PROCEDURE -- The following rules apply to single words or groups of words:

1. The first letter of each word is never deleted.
2. Deletion of letters commences from right to left, in order specified above.
3. Only one letter of a double letter occurrence is deleted.
4. Deletion is continued until the code word is reduced to a pre-determined size (number of letters).
5. Words already smaller than the predetermined size carry blank notations to complete the code.
6. Word is entered into its field left justified, hence blanks always occur on the right.
7. The size of the code word must be determined by study of the word population in question. Exceptions to the rules must be very few, but some cannot be avoided. A 5-letter code is satisfactory for most purposes.

Length of code word -- By this method, the length of a code word can be varied from one letter up to the length of the original word. The question of what specific length is "best", that is, the best compromise between uniqueness and brevity, can only be answered by a complete study of all geological names and terms, which study to the committee's knowledge is not available. Bourne (1963, Fig. 3-6) indicates that for a coding procedure very similar to that recommended here (Bourne's

No. 6), the uniqueness efficiency of a 5-letter code for "subject terms" is about 94 per cent -- that is, after coding to 5 letters, about 94 words out of 100 will be unique and the remainder duplicates. For 4 letters, the efficiency drops to about 87 per cent and it rises to nearly 100 per cent for 6 letters. In this connection, it is worth noting that there are other derived coding methods more efficient than the one recommended here, but they do not possess the same high mnemonic qualities.

There is very little virtue to shortness, as such, for a code word. Moreover, the longer the word the higher its degree of uniqueness and mnemonic qualities. Therefore, the committee recommends the use of a moderately long (5 or 6-letter) code for most purposes. The smaller the word population being coded, the shorter the code word length need be. Probably the best feature of this system is not with respect to any particular length of code word, but with the facility for establishing any desired maximum field size for a given population of words.

Examples of Recommended Coding Method

<u>Word</u>	<u>5-letter code</u>	<u>4-letter code</u>
Ireton	IRTON	IRTN
Jefferson	JFRSN	JFRS
Lloydminster	LDSMR	LDMR
Mesozoic	MSOZC	MSZC
Milk River	MLKRV	MKRV
Pennsylvanian	PNSLV	PSLV
Winnipegosis	WNPGS	WPGS
Buckinghorse	BCKGR	BCKG
Copper	COPPR	CPPR
Proterozoic	PRTZC	PTZC
Sandstone	SNDSN	SNDS
Archaean	ARCHN	ARCN
Old Fort Island	ODFID	ODFI

Conclusions

1. Heavily coded input/output formats are not necessary with modern computers, and coding at these stages should be minimized in the interests of accurate input and utility for the geologist.
2. An established derived, mnemonic coding method is recommended for the geological names and terms, where coding of these is necessary. As with all derived coding methods, some duplicates will appear which must be distinguished by some central authority if system-wide uniqueness is required.

3. The principal merits of the recommended method are:

- 1) all individuals and computers can independently generate the same codes,
- 2) the code has high mnemonic qualities, and
- 3) names and terms can be reduced to any desired word length. The main defect of the method is that complete uniqueness of code words generally is not possible for a given population of words.

Reference

1963, Bourne, C. P., Methods of information handling: John Wiley and Sons, New York, 241 p.

CHAPTER 8

NATIONAL INDEX TO GEOLOGICAL DATA

(F. Raynes)

The obvious need for a means of determining where data may be found has led to examination of the establishment of a National Index to Geological Data. Subcommittee 4, under chairmanship of F. Raynes, was assigned to a consideration of feasibility, that is, provision of answers to the following questions.

- (a) What are the basic requirements of a National Index?
- (b) How can these be achieved?
- (c) How should the Index be administered?

In addition to formal meetings, the Subcommittee conducted a workshop under the guidance of instructors from Imperial Oil Enterprises Limited to study the problems of indexing and to outline a preliminary policy for the National Index.

Requirements of a National Index

The utility of any data storage system depends upon selective retrieval from its contents. An index of these contents is the key to this selection and by varying techniques it defines the documents stored, indicating their subject content with the aim of leading users to data they are seeking.

A National Index to geological data is envisaged as an index to all published data and to those unpublished data on open file in Provincial and Federal government agencies, in universities and in other organizations which may volunteer to make data available to others; possibly, it could refer to data whose release may be conditional. It is desirable to try and index all available data regardless of the format of the particular file or document containing the data but unpublished data should receive priority. New data would form the bulk of indexed data; older data would be indexed depending on its validity and the economics of incorporating it into the system.

The significant characteristics of geological data files that might be referred to by a future National Index appear to be:

- (a) The source documents for the data would not be filed in a single location but would be filed in individual systems maintained by different agencies and organizations.
- (b) The individual data storage systems could employ a whole range of filing techniques, although the existence of a National System for data storage and retrieval using data processing equipment would, it is hoped, result in different organizations using this system in common.

It follows from these characteristics that the National Index system must be independent of the underlying data storage systems. As a consequence, a data retrieval search based on a National Index would identify rather than produce the relevant documents containing the data. Requests for unpublished documents would be made to the organization holding them; they would carry out a search of their files, forwarding the data either via original document or a copy, or would permit perusal on their premises according to the type of document and local policy. In the case of published data, the searcher, having found the title of the document, may discover he has his own copy, could purchase it or consult a library.

Index search for data will initially be by means of printed indexes. The potential bulk of indexed data would require the preparation of several indexes, each dealing with a specific aspect of geology. Additionally, the index system should be capable of producing special indexes on request. Later, as data processing system developments permit, index search may be done directly, that is, the system storage unit may be queried in a specific manner and the relevant index entries obtained as print-out.

Administration of a National Index

A National Index to Geological Data should be operated and maintained by a central agency. The central agency would be responsible only for receipt, processing and output of index material, and policy decisions must arise from a committee consisting of representatives of the organizations contributing to the Index. Standard index formats, as described below,

would be filled out and forwarded from the contributor (i.e. the owner of a document) to the central agency for conversion to machine processable form. This procedure will require a set of rules defining indexing policy to ensure the achievement of a common standard by various contributors. Such a central policy must be reviewed and possibly modified from time to time to reflect and reconcile the requirements of contributors and users and should take note of the evolution of the geological sciences.

An existing information system applicable to a National Index

Imperial Oil Enterprises Limited, in their Western Producing Region based in Calgary, have developed and put into operation an internal Technical Information Service using a coordinated indexing technique with a potential for adaptation in a National Index System. The system was developed to meet an increasing demand for information from their library facilities and their experience in terms of improved information recall and cost savings to date is reported to have justified the expense of installation.

Although the existing Imperial Oil System, known as the 'CAT' system, is a library indexing system the techniques employed appear to be equally applicable to a data index. Imperial Oil Limited have kindly offered to provide the relevant computer programs to the Geological Survey of Canada for use in a National Index System. The principles of the Imperial Oil System have been described by J.W. Cherry (1964, 1965).

In the context of its application in a National Index to Geological Data, the system requires the preparation of an Indexing Document (see Fig. 2) for each original document in the data files of the organizations contributing to the National Index. The skill with which Indexing Documents are prepared, in terms of the recognition and recording of the important contents of data documents, will determine the efficiency of the National Index in terms of future retrieval of relevant documents. Entries on the Indexing Document are transferred to punched cards forming the input to a data processing system programmed to examine, sort, and store on magnetic tape, each title, author, and concept.

As a matter of course, or upon request, indexes would be printed from these continuously updated tape files for distribution to users of the system.

Each index produced is in the form of an alphabetic listing of each concept that has been entered through the Indexing Documents processed in the system up to that time (see Fig. 3). The concepts recognized in all previously indexed documents appear as headings in alphabetic or numerical order on the left side of each page. Listed under each concept heading are the title and other bibliographic information referencing each document in which the concept has been recognized; these titles are arranged in ascending order of index serial number.

A search to retrieve data would begin with a definition of concepts that refer to the kind of data sought. These usually will include a location concept. Because the English language often provides many ways of defining the same basic concept the system must include a reference thesaurus listing any synonyms of the actual concepts appearing in the index. Armed with the approved concepts, the searcher would use these as coordinates to scan the index for the titles and owners of possible relevant documents. All relevant documents listed under each required concept would be compared using, if necessary, the index serial numbers to identify documents. If the same document number and title occur under each concept defined for the search, these will refer to documents that are most likely to satisfy the requirements of the investigator. Because the index shows the title of each document, which is often an additional clue to the detailed content of the document, it is frequently unnecessary to examine every concept defined to conclude the search; titles frequently may enhance the relevancy of retrieval.

A coordinate search does not necessarily ensure relevance. It is obvious that success depends on the skill with which the document has been indexed and the depth of the indexing.

Indexers must be properly trained, have access to a clear set of guidelines and be capable of recognizing the essential data content of a document. They must, in effect, predict what future users may want from any document indexed. An indexer must be familiar with the requirements of the geological sciences and, ideally, would be a competent geologist himself. The task of indexing of concepts cannot be delegated to inexperienced junior staff as a kind of clerical function.

The Indexing Document used by Imperial Oil, and being employed in the pilot indexing project of the Geological Survey, is shown in Figure 2,

The information collected falls into the following categories:

- (a) The serial number of the Index entry.
- (b) The title and other bibliographic data sufficient to identify a data document, where it is stored and the organization maintaining this storage system. The local or owner's reference number of the document would form part of this information (see Appendix VII, part A).
- (c) The type of Output Control to be applied in the preparation of printed indexes. This defines which indexes will contain each index entry derived from Indexing Documents. For example, output control would permit the preparation of an index to palaeontological data or other specified geological discipline, or an index to the data held in the files of the Ontario Department of Mines, or an index to geological maps of Canada, and so on (see Appendix VII, part B).
- (d) The type of document containing the data (see Appendix VII, part C).

(e) The concepts that describe the geographic location and kind of data (see Appendix VII, parts D & E). For example, a report on a geochemical survey may result in the indexing of the following concepts:

Nova Scotia
11-K
Soil Geochemistry
Stream Silt Geochemistry
Copper
Lead
Zinc
Cold Extraction
Dithizone
Sampling Techniques

This set of concepts would have been recognized by an indexer if, for instance, the document in question contained data in the form of a set of tables and a map recording the results of soil and stream silt geochemical sampling and testing for cold extractable copper, lead and zinc using dithizone to indicate the presence and amounts of these heavy metals. The same document would have referred to sampling techniques and the sampling would have taken place in Nova Scotia, in particular in an area within the National Topographic System lettered quadrangle 11-K.

There are practical limits to the depth or completeness of indexing and the degree of specificity of a concept. Specificity in information retrieval parlance refers to the degree to which a concept recognizes an exact "species" rather than the containing "genus". Exhaustivity depends on the policy adopted for recording concepts on the Index Document and the extent to which the indexer recognizes these concepts. Specificity and exhaustivity apply as well in the search of the index and the definition of concepts that are too narrow or specific in meaning may result in no retrieval of useful data. Similarly, the definition of too many concepts in a search reduces the probability that any document will be found under all of them.

No recommendations regarding the establishment of a list of authorized concepts are possible until specific disciplines of geological sciences are considered for indexing. In the Imperial Oil System, the approved list of concepts is flexible and may be altered and lengthened as required. Its use is enhanced by the thesaurus to convert synonyms to approved terms and, for concepts of low specificity, indicates the generic term to be used when more exact but rarely used concepts do not appear on the list.

A pilot study to reveal the real problems of establishing a National Index using the recommended system was started with the National Index workshop in May 1966. This study will continue until September and will index certain published and unpublished Geological Survey of Canada

documents and develop a Concept Authority list and examples of a final printed index. A final design for the Index system, the setting up of a policy framework and outlining a possible organization to maintain the Index will be considered in the light of this experience.

Recommendations and conclusions

The subcommittee's conclusions on the desirability and feasibility of a National Index to Geological Data are summarized as follows:

- (a) A National Index is highly desirable and appears feasible.
- (b) A central agency with adequate staff would be required both to house and maintain the Index. It is suggested that the Geological Survey of Canada assume the responsibilities of this agency in the formative stages of the Index. The detailed organization of the ultimate agency and its position within the structure of Federal Government services remain to be investigated. Close cooperation between the contributors to the Index would be essential and a central National Index policy committee would have to be established at an early stage.
- (c) The Index should be based on the system designed and currently used by Imperial Oil Enterprises Limited (Western Producing Division).
- (d) The Index can and should deal with the bulk of geological data currently being produced and subsequently should be extended to include that produced in the past which is deemed to be of current or future interest.
- (e) To provide a demonstration of the system for the National Advisory Committee and potential users and contributors, a pilot study is being conducted in Ottawa indexing various files of the Geological Survey of Canada. The results of this pilot study are expected to provide a basis for policy decisions on indexing procedures, as well as observation of, and provision of remedies for, any weaknesses in the subcommittee's initial ideas and recommendations.

References

- 1. J.W. Cherry "Automation and Information Retrieval", Proceedings of the Computing and Data Processing Society of Canada, 4th Conference, Ottawa, May 1964.
- 2. J.W. Cherry "A Computer-Assisted, Industry-Oriented, Information Retrieval System", Technical Information Services Department, Imperial Oil Limited, (Producing Department, Western Region), April 1965.

CHAPTER 9

DATA ON MINERAL DEPOSITS

Introduction

In a review of geological subjects for which machine-processable data files are desirable, the field of mineral deposits was among the most prominent. Accordingly Subcommittee No. 5, under the chairmanship of D.R.E. Whitmore was appointed to determine what data on mineral deposits should be recorded, what terminology should be used, and to develop a format for recording the data.

Files of data on mineral deposits have been maintained by Provincial Departments of Mines dating back to the 19th century and most large mining and exploration companies have voluminous records of properties that they have investigated. Additional files have been compiled by the Federal government under the stimulus of war-time conditions. Many of the data accumulated by Provincial authorities, (and by the Federal authorities in the Northwest and Yukon Territories) have been published in annual reports, reviews, etc.

Existing files have been used primarily for reference to individual deposits and secondarily for compilation of technical data, estimates of reserves, etc., and their formats were designed accordingly. The proposed machine-processable format must serve these purposes more effectively than its predecessors, and must also make possible statistical evaluation of geological data related to ore deposition. If this can be accomplished the system is certain to find wide use by industry and in other fields of geological research.

The principal problems to be reviewed by the committee are:

1. Definition of a mineral deposit for purposes of this file.
2. Geological and other data that are essential for all mineral deposits.
3. A level of observation and terminology that will ensure consistency in use of the file.

If these geological factors can be satisfactorily decided and combined with the four general factors that apply to all subject files (see Chapter 3), it should be possible to design a format that could serve most applications,

Although it is recognized that for modern files, punched cards serve mainly as input to magnetic tape, discs, or other forms of storage, it is convenient to record the proposed format in terms of cards so that provisions for extension of data for special purposes can be readily visualized.

Uses to be made of the file

Any record of a mineral deposit by a government or a mining company will be used primarily for reference. Many of the data required for reference purposes will not be geological; they will consist of location, means of access, status of mining claims, ownership, development work, assays etc. and it is likely that in many files reference to such data will be more frequent than reference to geological facts.

Geological data are used primarily to assess the economic potential of each individual deposit. In recent years such assessments have been materially assisted where direct comparisons of data of the deposit under review could be made with data of geologically similar deposits elsewhere. These comparisons have led to development of empirical criteria that may be applied directly to the facts of any single deposit. Such criteria can be tested, modified and made more effective by use of files of objectively-recorded data for many deposits. Statistical evaluation of data of mineral deposits is being developed to classify deposits and to test hypotheses concerning broad geological associations in time and space, provenance, and processes leading to concentration of metals in orebodies. This use has great potential in both scientific and practical applications.

All these purposes would benefit by establishment of files of data from large numbers of mineral deposits consistently recorded in a format compatible with computer processing.

Existing files

Although machine-processable files have been developed for data related to petroleum deposits in Canada there are as yet no such files for data of mineral deposits. In recent years however, mining companies, and particularly Provincial Government departments, have recognized the importance of obtaining the same data for all deposits and for this purpose have devised specific forms on which the data are recorded. Most such forms are printed on large cards to facilitate sorting and some use edge-punched cards. The main advantage in changing from these to a machine-processable format is the computer's ability to sort by location, metal, mineral, or any other of the factors recorded; to compare data; to make statistical analyses, etc.; and to report results automatically in tabular or graphic form.

The subcommittee, through the courtesy of the organizations concerned, has had the opportunity to study formats used by the following:

1. New Brunswick Department of Lands & Mines
2. Quebec Department of Natural Resources
3. Ontario Department of Mines

4. Manitoba Department of Mines & Natural Resources
5. Research Council of Alberta
6. British Columbia Department of Mines & Petroleum Resources
7. Union Carbide Explorations Limited
8. Mineral Resources Division, Department of Mines and Technical Surveys
9. Geological Survey of Canada

These formats in large measure reflect the purposes for which they were designed. As would be expected many of them provide space for data required by provincial mining regulations on the legal status of claims, development, and other technical data. Because they are used largely for evaluation of the production potential of each deposit, it is usual to record conclusions and assessments of the observer in addition to the facts.

Review of these formats indicates that their contents can be classified in four main categories:

1. Reference, development and location data
2. Geological data
3. Geological and economic conclusions
4. References, observer's name, maps, reports etc.

Common entries are:

- 1a. Reference number
- b. Name of deposit
- c. Name and addresses of present and past companies
- d. Claim numbers, type, total area etc.
- e. Status of claims
- f. Location (i) of property
(ii) of nearest shipping point
- g. History of exploration
- h. Present development
- 2a. Metal or mineral sought (i) major
(ii) minor
- b. Form of deposit (vein, massive, disseminated etc.)
- c. Shape and size of deposit
- d. Mineralogy (i) metallic
(ii) gangue
- e. Associated rocks (i) host rocks
(ii) wall rocks
(iii) other adjacent rocks
- f. Structure (i) regional
(ii) local at deposit

- g. Age (i) of enclosing rock
 (ii) of other rock
 (iii) of ore
- h. Samples and assay results
- i. Other geochemical data
- j. Geophysical data
- 3. Conclusions (a) Principal ore control
 (b) Provenance of ore
 (c) Genetic process and association
 (d) Geological classification of deposit
 (d) Economic assessment
- 4a. Observer's or compiler's name
- b. Unpublished company reports of assessment work etc.
- c. Published reports referring to deposit (Minister of Mines Reports
etc.)

Definition of a mineral deposit

For purposes of the file it is necessary to define what is to be considered a mineral deposit. Ideally a mineral deposit should be a discrete, continuous entity with geologically definable limits of length, width and ultimately, depth. Such a situation is rarely met and more commonly a deposit consists of several bodies occurring in parallel structures, often en echelon and which are usually similar in mineralogy, elemental content, etc. In such cases, it may be necessary to select and describe a typical ore body.

In practice it is particularly difficult to define a mineral deposit, until a large amount of development work has been done. There is the further complication that two or more companies may mine different parts of a single mineral deposit and conversely that one company may mine two or more different mineral deposits from a single shaft.

The element of geological time is a further complicating factor. Many deposits contain evidence that suggests that the ore was modified subsequent to its original deposition. In uranium deposits of the Beaverlodge area for example, generations of pitchblende whose age differs by 1000 million years are found in the same vein. Such composite mineral deposits are likely to complicate use of the file for comparative purposes.

The subcommittee has recognized these various problems but has not formulated a definition of a mineral deposit as the 'unit of record' for this file, pending further consideration of existing practice and of the purposes to be served by the file.

Development of a format for the National System

For purposes of the proposed National System, the main concern is with geological data of mineral deposits and development of the file is concerned primarily with selection of data that should be recorded, and secondarily with the level of observation at which the data can be recorded consistently.

However the file must also serve other purposes for which files exist and provision must be made for parallel records of technical, legal and other data as part of, or suitably cross-referenced to, the file of geological data. In this respect, development of data files on mineral deposits will be analogous to development of files on oil wells and review of the latter will certainly contribute to discussions of the subcommittee.

Success of the file will depend primarily on the cooperation of the organizations and geologists who will observe and record the data. The file format must be as self-evident as is possible, and must be designed so as to reduce ambiguities to a minimum and at the same time minimize perusal of long explanatory documents, codes etc.

Undoubtedly a format that must meet so many requirements will be a compromise and will have to develop by trial and error. The subcommittee has adopted the sound policy of developing a machine-processable file by examining existing files and modifying them to fulfil the geological purposes of the National System. Any format that is developed will have to be tested in the field and further modified as experience dictates. Moreover, the format will have to be flexible enough to provide for records of additional data and even conclusions that are needed by the various cooperating organizations for their own purposes. Provision will have to be made however to distinguish the data required by the National System from the data and information required for other purposes so that there will be consistency in the data available in the files of all those organizations that choose to make use of the national format.

In order to discharge its responsibilities the subcommittee has devoted its efforts toward the development of a standard format for recording geological data and is also indicating those other data necessary for reference or indexing.

Level of observation

Chapter 6 of this report reviews factors involved in recording descriptive data so that in the terms used there will be universal understanding and freedom from conclusions by the observer. This is of particular importance in recording data concerning mineral deposits, a field in which many terms have genetic connotations. Probably it will be

impossible to avoid use of some of these terms but the following simple example may indicate why records of fact are essential in archival files for general use.

In the walls of several major deposits in the Bathurst area of New Brunswick, a rock occurs that is recognizable principally by disseminated grains of quartz larger than the grain size of the matrix. In different deposits this rock has been recorded as 'ignimbrite', 'quartz porphyry', and 'quartz-eye sediment'. A somewhat similar rock occurs in the Flin Flon mine as part of a mapping unit generally referred to as 'the sericite schist'. Each of these names is well-recognized and understood for reference purposes in the mine where it is used. However, if such rock names are recorded as data in an archival file, it is most unlikely that similarities in the actual rocks would be recognized and any significance there may be in their association with the mineral deposits would be lost.

Clearly it is essential in recording data of mineral deposits, that observations be made at a level of sophistication low enough and in terms simple enough to ensure that the facts are clearly understood by the users of the file. Where complex terms must be used, they should be defined. Consistency will be improved by asking the observer to check one of a limited number of terms. In those parts of a form where the observer is left to provide his own terms, a list of suggested terms or phrases often assists the observer to comprehend the type of data required.

Progress of subcommittee

There is a wide disparity in the formats of existing files of data on mineral deposits and the subcommittee had to decide where its responsibilities lay. It was decided that decisions on geological data that should be included comprised its primary responsibility but that decisions on records of other data must also be considered. Two schools of thought emerged: one group advocated a minimal factual record supplemented by reference to reports and the other recommended extensive records of data and possibly some information. Table 3 shows a compromise format that is currently under discussion. It will be noted that the first two pages record index and reference data and that the remaining 4 pages record geological data describing the deposit and its environment. The remarks column is used to amplify the main headings as needed. Despite these precautions it will be necessary to issue a pamphlet or manual to explain how the form is to be filled out and possibly to define certain terms. Some consideration is being given to printing a sample of a filled form on the back of each sheet. It is likely the six sheets will be combined into a single folded sheet.

CHAPTER 10

DATA ON FOSSIL FUELS AND RELATED ECONOMIC SEDIMENTARY DEPOSITS

(C.F. Burk, Jr.)

Introduction

A file of geological data concerning fossil fuels and other related economic sedimentary deposits was considered by the ad hoc Committee as an appropriate example of a particular data file that could be developed within the framework of general proposals outlined by the committee, and a beginning was made to build this file. Geological data in this area are generally available in the public domain, particularly those pertaining to oil and gas deposits; the volume of data is reasonably large; and there appears to be interest by government and industry in having such data available in machine-processible form.

The essential questions faced by the committee were:

- (1) What geological parameters or descriptors are necessary to adequately describe a fossil fuels deposit?
- (2) What standards of measurement and description should be applied to the selected descriptors?

New generations of computing hardware and improvements in the software technology have removed most of the problems of file design. The aspects of basic importance to the creation of a successful public file are the selection of input entries adequate to the demands expected of the output, and the establishment and acceptance of standards for the input entries.

The Committee has made some progress in relation to these problems. There remain the preparation of a detailed specifications manual, writing of a retrieval program, testing the adequacy of the input, and testing the retrieval program. Finally, the usefulness and significance of the proposed file must be evaluated.

Purpose of proposed file

The purpose in recommending standards for establishing a fossils fuels data file is to make more easily and widely available, in a standardized form, fundamental geological data on these deposits.

With the establishment and acceptance of standards and universal availability of the file design and programs, there is no fundamental reason

to have a single centralized file to act as sole authority, nor for any file to be restrained in size or scope.

The availability of comprehensive computer files of geological data on fossil fuels and related deposits will allow the user to retrieve data efficiently on a particular deposit in which he is interested, and will permit an analysis of the entire file that could serve as a basis for geological research into the origin and habitat of the deposits.

An example of one approach to studying fossil fuels deposits that this file would make possible for Canadian deposits, is provided by Knebel and Rodriguez-Eraso (1956), who studied 236 major oil fields of the "free world". These fields were analyzed with respect to location, basin position, type of trap, lithology of reservoir, gravity of crude, depth of occurrence, and time of discovery.

Perhaps the most important purpose which would be served by the proposed file is that of having a current, immediately accessible inventory of fossil fuels deposits in terms of their geology, which could form a substantial and comprehensive basis for policy decisions by both government and industry in connection with such areas as exploration, development, administration, and economics.

Unit of description

The basic unit with which the file deals is the deposit, which may be considered, conceptually at least, as a three-dimensional body of strata containing oil, gas, coal, or other materials currently of economic value, in sufficient abundance that they are being, or are capable of being, economically exploited. From a broad geological point of view, the limits of such deposits are arbitrarily defined, depending as they do on such factors as current economic value and current technology. Nevertheless, the delineation of a particular deposit at a particular time depends on the recognition and mapping of certain geological variables.

It is the committee's intention, as an operational procedure, to deal only with known economic deposits and to exclude from consideration, for the time being, data on other deposits that are not now economic. The status of deposits in this respect fluctuates with time and clear lines of demarcation are lacking in some cases; such borderline cases will be considered on their individual merits, with the hope that eventually all deposits will be filed, including those that could be classed as "potentially" and "formerly" economic.

In connection with oil and gas deposits, it should be clearly understood that the file proposed here is not a "well data file", many of which are currently in operation (e.g. Buller, 1964). The basic unit described in such

files is the well or borehole, whereas the basic unit considered here is the deposit.

General approach

The Committee approached the question of selecting appropriate data entries by initially considering the problem from an abstract or theoretical point of view, and subsequently examining the types and sources of data actually available. Data entries were then modified accordingly. The result is the list of data entries and general information shown in Table 4. Further minor modifications are expected when detailed specifications are considered and test projects are completed. However, the Committee considered its prime responsibility to be the selection of theoretically appropriate geological descriptors and the establishment of standards, and secondarily, questions related to data availability and a modus operandi.

The level of sophistication was kept relatively low in order to ensure a high degree of consistency. As previously mentioned, consistency is a fundamental requirement to any successful data file. A general rule in this respect seems to be: the greater the number of expected contributors and sources of data to a file, the lower the level of sophistication required to maintain satisfactory consistency.

Nature of data

As discussed elsewhere in this report (see Chapter 1), the ad hoc Committee has directed its attention to geological data, and not to information. It is recognized, however, that a clear distinction between the two is not easily achieved. Among other things, the distinction depends on the level or point of view from which the distinction is attempted -- one man's information is another man's data, and vice versa.

From the point of view of a file on deposits, data comprise those measurements and descriptors that characterize the deposit as a whole. Thus, these data cannot be used to analyze individual deposits; they can be used, however, to compare one deposit as a whole with another. It happens that knowledge about most deposits is obtained from a finite number of intersections with the deposit (wells, shafts, drill holes, etc.). Although from a well-data compiler's point of view, data are defined in terms of measurements made in individual boreholes, from the point of view taken here data are represented by items derived from the processing of many borehole data, which items can be said to apply objectively to the deposit as a whole. The question arises as to how the borehole data should be processed to insure consistency in results obtained from multiple sources. It was apparent to the committee that agreement on standards for computation of some items (e.g. average effective porosity for a hydrocarbon pool) does not

exist at present; for most items, more or less standard procedures have evolved and have been accepted. As the file develops, increasing attention should be focused on standardizing these procedures. For the present file, the effect of different processing procedures has been minimized by the generally low-order sophistication of data, the purposeful avoidance of particularly difficult items of data (e.g. permeability), and provision for such entries as "variable" and "unknown".

The recommended data entries (Table 4) were broadly classified in four categories:

1. Index data -- e.g. name, location.
2. Historical data -- e.g. discovery date.
3. Geological data -- e.g. lithology, geologic age.
4. Technological data -- e.g. number of wells.

Emphasis in the file is of course on category 3 (geological data). Some technological data are included to provide a broad indication of the methods and scope of the extraction technology, but there is no intention to give comprehensive treatment to this category.

The geological data category can be further broken down into three classes with respect to the commodities being extracted from the deposit:

1. Data on the commodities per se.
2. Data on the geology of the deposit.
3. Data on the geologic setting of the deposit.

Data entries and general specifications

Table 4 lists the committee's recommended data entries and general specifications. These recommendations, as stated at the outset of this chapter, are by no means final, and should be regarded as provisional. It is hoped that through publication, wider interest and concern in this area will be stimulated and practical test programs conducted, all of which will assist in the eventual building of a satisfactory and useful file.

The emphasis of Table 4 is on hydrocarbon deposits. This reflects the relatively high current value of these deposits, the generally wide availability of data on them, and, not least, the professional background of the committee members. Nevertheless, the recommended geological descriptors are equally applicable to other classes of deposits such as coal and evaporites. Specialized detail appropriate to these and other deposits remains to be investigated by subsequent committees. Table 5 provides a suggested numeric hierarchical code for geologic time units.

Future work on file development

The following steps appear to be necessary in order to establish a usable fossil fuels data file.

1. Preparation of detailed specifications manual:

The general specifications of Table 1 need to be considered and refined in detail to ensure consistency of data input. Definitions and procedures must be explicit and unambiguous.

2. Design of input format:

Based on the detailed specifications manual, an input format must be designed. This should be done in cooperation with a systems analyst to ensure maximum efficiency for retrieval programs.

3. Writing of a computer retrieval program:

4. Testing of adequacy of data input specifications and retrieval program:

5. Evaluation of file:

Ultimately, only the users of the file can properly evaluate the usefulness or otherwise of the data. Provision must be made for continual improvement of the file content and the quality of data collected.

REFERENCES

Buller, J.V.

- 1964: A computer-oriented system for the storage and retrieval of well information; Bull. Can. Petrol. Geol., vol. 12, no. 4, pp. 847-891.

Knebel, G.M. and Rodriguez-Eraso, G.

- 1956: Habitat of some oil; Bull. Am. Assoc. Petrol. Geol., vol. 40, no. 4, pp. 547-561.

Levorsen, A.I.

- 1954: Geology of petroleum; San Francisco, W.H. Freeman and Co., 703 p.

MacKay, B.R.

- 1947: Coal reserves of Canada; reprint of Chapter 1, Appendix A, Rept. of Royal Commission on Coal, 1946; 113 p.

APPENDIX I

MEMBERS OF THE SUBCOMMITTEES

Subcommittee No. 1

Reference numbering and geographic location

Mr. D.A. Sharp, Chairman
Energy Branch,
Department of Energy and Resources
Management,
880 Bay Street,
Toronto 5, Ontario.

Members

Dr. B.K. Bhattacharyya,
Geological Survey of Canada,
601 Booth Street,
Ottawa 4, Ontario.

Mr. A.E. Buller,
Manager, Union Carbide Exploration Limited,
805 Davenport Road,
Toronto 4, Ontario.

Mr. J. Foster,
Central Data Processing Service Bureau,
Government of Canada,
Ottawa, Ontario.

Mr. J.A.M. Haddon,
Chief Cartographer.
Ontario Department of Mines,
Toronto, Ontario.

Mr. C.E. Hoganson,
Chief, Computing and Control,
Surveys and Mapping Branch,
615 Booth Street,
Ottawa 4, Ontario.

Dr. S.S. Holland,
Mineralogical Branch,
B.C. Department of Mines and
Petroleum Resources,
Victoria, B.C.

Mr. R.J. MacGregor,
British American Oil Company,
Calgary, Alberta.

Dr. D.L. Stauff,
Imperial Oil Enterprises,
500 Sixth Avenue South West,
Calgary, Alberta.

Subcommittee No. 2

Reference Numbering and Geographic Location

Dr. S.C. Robinson, Chairman
Chief, Economic Geology Division,
601 Booth Street,
Ottawa 4, Ontario.

Members

Dr. Robt. Assad,
Mineral Deposits Service,
Department of Natural Resources,
Quebec City, P.Q.

Dr. K.R. Dawson,
Geological Survey of Canada,
601 Booth Street,
Ottawa 4, Ontario.

Mr. N.M. Ediger,
Economic Geologist,
The British American Oil Co. Ltd.,
P.O. Box 130,
Calgary, Alberta.

Dr. J.E. Reesor,
Geological Survey of Canada,
601 Booth Street,
Ottawa 4, Ontario.

Dr. D.R.E. Whitmore,
Geology of Mineral Deposits,
Geological Survey of Canada,
601 Booth Street,
Ottawa 4, Ontario.

Dr. Derek M.R. Wilson,
Northern Alberta Institute of Technology,
106 Avenue and Princess Elizabeth Avenue,
Edmonton, Alberta.

Subcommittee No. 3

Mnemonic codes

Dr. C.F. Burk, Jr., Chairman
Geological Survey of Canada,
406 Customs Building,
Calgary, Alberta.

Members

Dr. Robt. Assad,
Mineral Deposits Service,
Department of Natural Resources,
Quebec City, P.Q.

Mr. J.W. Cherry,
Imperial Oil Enterprises,
500 Sixth Avenue S.W.,
Calgary, Alberta.

Subcommittee No. 4

National Index to Geological Data

Mr. F. Raynes, Chairman
Senior Mining Analyst,
Exploration Division,
Cominco Limited,
P.O. Box 1510, Station "B",
Montreal, Quebec.

Members

Dr. Robt. Assad,
Mineral Deposits Service,
Department of Natural Resources,
Quebec City, P.Q.

Mr. J.W. Cherry,
Imperial Oil Enterprises Limited,
500 Sixth Avenue S.W.,
Calgary, Alberta.

Mr. A.S. Gibson,
Geophysicist,
Imperial Enterprises Limited,
500 Sixth Avenue S.W.,
Calgary, Alberta.

Mr. J.A.F. Haddon,
Ontario Department of Mines,
Parliament Buildings,
Toronto, Ontario.

Dr. Stuart Holland,
Mineralogical Branch,
B.C. Department of Mines and
Petroleum Resources,
Victoria, B.C.

Dr. J.R. Smith,
Head, Geological Division,
Saskatchewan Research Council,
Saskatoon, Saskatchewan.

Dr. D.R.E. Whitmore,
Head, Geology of Mineral Deposits,
Geological Survey of Canada,
601 Booth Street,
Ottawa 4, Ontario.

Dr. K.R. Dawson,
Geological Survey of Canada,
601 Booth Street,
Ottawa 4, Ontario.

Subcommittee No. 5

Storage and retrieval of data on mineral deposits

Dr. D.R.E. Whitmore, Chairman
Head, Geology of Mineral Deposits Section,
Geological Survey of Canada,
601 Booth Street,
Ottawa 4, Ontario.

Members

Dr. R. Assad,
Director, Mineral Deposits Service,
Department of Natural Resources,
Quebec, P.Q.

Prof. W.C. Brisbin,
Department of Geology and Mineralogy,
University of Manitoba,
Winnipeg, Manitoba.

Mr. A.E. Buller,
Manager, Union Carbide Explorations Ltd.,
805 Davenport Road,
Toronto 4, Ontario.

Dr. R.L. Cheesman,
Chief Geologist,
Precambrian Geology Division,
Department of Natural Resources,
Regina, Saskatchewan.

Dr. J.F. Davies,
Chief Geologist,
Department of Mines and Natural Resources,
Norquay Building, 401 York Avenue,
Winnipeg 1, Manitoba.

Mr. W.D. Harrison,
Geologist, Fecunis Division,
Falconbridge Nickel Mines Ltd.,
Onaping, Ontario.

Dr. D.F. Hewitt,
Senior Geologist, Department of Mines,
Parliament Buildings,
Toronto, Ontario.

Dr. S.G. Ferguson,
Geologist, Department of Mines,
Parliament Buildings,
Toronto, Ontario.

Dr. S.S. Holland,
Geologist, Department of Mines and
Petroleum Resources,
Victoria, B.C.

Mr. R.R. Potter,
Geologist, Mines Branch,
Department of Lands and Mines,
P.O. Box 758,
Fredericton, N.B.

Subcommittee No. 6

Fossil fuels and related economic sedimentary deposits

Dr. C.F. Burk, Jr., Chairman
Geological Survey of Canada,
406 Customs Building,
Calgary, Alberta.

Members

Dr. J.V. Buller,
Saskatchewan Department of
Mineral Resources,
Regina, Saskatchewan.

Dr. S.S. Cosburn,
British Columbia Department of Mines
and Petroleum Resources,
Victoria, B.C.

Mr. N.M. Ediger,
British American Oil Company Limited,
Calgary, Alberta.

Mr. L.A. Falkenberg,
Alberta Oil and Gas Conservation Board,
603 Sixth Avenue S.W.,
Calgary, Alberta.

Dr. M. Houde,
Quebec Department of Natural Resources,
Quebec, P.Q.

Mr. S. Kanik,
Department of Northern Affairs and
National Resources,
Ottawa, Ontario.

Dr. H.R. McCabe,
Manitoba Department of Mines and
Natural Resources,
Winnipeg, Manitoba.

Mr. D.A. Sharp,
Ontario Department of Energy and
Resources Management,
880 Bay Street,
Toronto 5, Ontario.

Mr. D.L. Stauff,
Imperial Oil Enterprises Limited,
500 Sixth Avenue S.W.,
Calgary, Alberta.

Subcommittee No. 7

Data of Geophysics

Mr. A.S. Gibson, Chairman
Imperial Oil Enterprises Limited,
500 Sixth Avenue S.W.,
Calgary, Alberta.

Member

Dr. B.K. Bhattacharyya,
Geological Survey of Canada,
601 Booth Street,
Ottawa 4, Ontario.

Other members to be co-opted.

Subcommittee No. 8

Survey of Machine Processing in the Geological Sciences

Mr. N.M. Ediger,
The British Americal Oil Co. Ltd.,
P.O. Box 130,
Calgary, Alberta.

Other members to be co-opted.

APPENDIX II

ADVANTAGES AND DISADVANTAGES FOR DATA PROCESSING OF THE VARIOUS COORDINATES USED FOR GEOGRAPHIC LOCATION IN CANADA

1. Geographic coordinates (latitude and longitude)

It is assumed that these geographic or angular coordinates are so well-known that description is unnecessary. They are rarely printed as a grid on maps but appear in a graticule around the edge of maps with some intersections, plotted as crosses, in the body of the maps. Neatlines (inner margins) of 1:250,000 maps are divided into degrees and units of 15 minutes whereas neatlines of 1:25,000 maps are in degrees and units of 5 minutes.

Advantages:

- (a) Universally applicable without discontinuities.
- (b) Excellent for computation of distance, area, and volume over large regions.
- (c) Available on maps of all scales.
- (d) Widely used and understood.
- (e) Computer programs available for conversion to U.T.M.

Disadvantages:

- (a) Grid constructed of converging ordinates and curved abscissae.
- (b) Unsatisfactory for reference due to difficulty of precise measurement of coordinates on maps.
- (c) Variation in linear distance per degree with difference in latitude.
- (d) Require use of computer to establish rectangular coordinates for plotting by mechanical devices.
- (e) Degrees, minutes and seconds are not a simple decimal system. Decimals of degrees are not shown on maps.

2. The National Topographic Series grid

This grid consists of numbered areas bounded by meridians and parallels of latitude with dimensions of 4 degrees of latitude by 8 degrees of longitude. Each area is subdivided into 16 lettered areas each 1 degree of latitude by 2 degrees of longitude in size. Each of these is further subdivided into 16 numbered areas each 15 minutes of latitude by 30 minutes of longitude. Finally there is a further subdivision into 8 lettered areas each 7 1/2 minutes of latitude by 7 1/2 minutes of longitude.

This system is used only in Canada and is designed to identify map-areas. If used for point location, a secondary grid must be introduced. This reduces usefulness for computation and results in coordinates expressed in two different units; angular and linear. Use of a secondary grid results in

discontinuities at every degree of latitude and longitude, at least. For point reference and computation therefore, this system is inferior to the geographic coordinates on which it is based.

3. The township and range system

Legal surveys of most provinces are based on a system of townships in ranges whose boundaries are roughly north-south. This system is based on geographic coordinates with adjustments made as necessary along the east-west boundaries. In settled areas, range and township lines are surveyed and even subdivision and lot lines or their intersections may be marked. This system of coordinates is discontinuous, covers only a small part of Canada, does not permit definition of a point without introduction of a secondary grid, varies from province to province and is markedly inferior to geographic coordinates for scientific purposes. Nevertheless it is the legal basis for location of oil wells and water wells in some provinces and provision may have to be made to record location of such wells in terms of this system.

4. Marsden squares

This is a world-wide system of numbered areas each 10 degrees of latitude by 10 degrees of longitude, that has been used internationally for oceanographic and meteorological reference. These squares are numbered sequentially in tiers around the equator, starting at 0 degrees longitude and increasing to the west. Thus the square between 0° and 10° West and 0° and 10° North is No. 1, that between 10° and 20° West and 0° and 10° North is No. 2 and that between 10° East and 0° East between 0° and 10° North is 36. Squares 37-72 form the 2nd tier and so on up to 80°N. The sequence is resumed south of the equator. Provision is made for subdivision of each 'square' into 100 smaller areas each 1 degree by 1 degree but for point location a secondary grid would be needed with resulting discontinuities and use of two sets of coordinates; one angular, the other linear. Modern trends appear to be away from this system. It would not be suitable for location of geological samples. Marsden squares appear on a map of Canada at a scale of 100 miles = 1 inch, No. S 161 published by the Hydrographic Service of the Surveys and Mapping Branch, Mines and Technical Surveys, dated 20 January, 1961.

5. The Universal Transverse Mercator grid

This is a world-wide system of zones, each zone being 6 degrees longitude wide and each extending from the equator to the 80th parallel of latitude, north and south. In Canada the grid is extended 84°N. Within each of the 60 zones there is a rectangular grid in metres. The ordinates of this grid are parallel to the central meridian of each zone and the abscissae are normal to it. Thus all squares (1,000 metre on 1:250,000 maps) are the same size throughout each zone. Northings measured in metres extend from 0 at the equator to slightly over 9,200,000 at 84°N; and eastings, also in metres, are measured from the central meridian which is arbitrarily assigned an easting of 500,000 metres. At the 49th parallel eastings extend from

approximately 280,000 at the western margin of a zone to approximately 720,000 at the eastern margin, whereas at the 84th parallel the extent of eastings is only from about 470,000 to 530,000. Although the abscissae of adjacent zones meet at the zone boundaries there is an angular discontinuity and hence the grids are continuous only within each zone. This may be seen in maps crossing the boundaries such as 97C, the Franklin Bay map-sheet.

U.T.M. grids are available on all topographic maps at scales of 1:250,000, 1:50,000 and 1:25,000 and on some maps at 1 inch to 1 mile, 1 inch to 4 miles and 1 inch to 8 miles. Maps of Canada at a scale of 1:2,000,000 are also available with the U.T.M. grid.

Advantages

- (a) A uniform rectangular grid.
- (b) Excellent for reference; coordinates in metres are readily measured on maps and may be used directly on mechanical plotting instruments.
- (c) Available on maps of United States and Canada at scales of 1:250,000, 1:50,000 and 1:25,000. Applicable to both Mercator and Polyconic Projections. The official grid of free-world armies.
- (d) Excellent for computations of areas within each 6-degree zone.
- (e) Computer programs and tables are available for conversion to latitude and longitude.

Disadvantages

- (a) Grid discontinuities at zone boundaries (zones are 6° of longitude wide and extend from equator to 80°N and S.) in areas extending over more than one zone.
- (b) Little known due to relatively recent introduction on maps for civilian use.
- (c) Measurement in metric system.
- (d) Restriction to maps of larger scales only.

In order to ascertain the relative accuracy and speed with which a point on a map can be measured or read off, in terms of this U.T.M. grid compared to latitude and longitude, a study was made by R. Langlois, F. Zieman and Y. Roy of the Geophysics Division, Geological Survey of Canada. Their report is as follows:

Materials and equipment used

- (1) Maps
 - (a) 1 sheet in the scale of 1:25,000;
 - (b) 1 sheet in the scale of 1:50,000;
 - (c) 1 sheet in the scale of 1:125,000.
- (2) A Gerber Variable Scale with an accuracy of 0.05% in 20 inches.
- (3) Half-inch square grid overlay with 660 points.

Preparation of maps

In order to determine on a map the location of a point in degrees, minutes and seconds with reasonable accuracy, it is necessary to draw grid lines spaced at every minute or 30 seconds. The length of time taken to draw these lines varies with the required accuracy and the type of map being studied. The following table gives the times required to grid completely different types of maps:

<u>Map Scale</u>	<u>Grid spacing</u>	<u>Time required in minutes</u>
1:25,000	30 seconds	20 - 30
1:50,000	60 seconds	40 - 50
1:250,000	60 seconds	150 - 180

Grid lines are always printed on maps with Universal Transverse Mercator projection with a spacing of 1000 metres when the scale is 1:25,000 or 1:50,000 and with that of 10,000 for the scale of 1:250,000.

Reading and plotting the grid points

The time needed to read locations of a set of points from a map and to plot those back on a map depends on various factors such as the complexity of the map, the coordinate system used and the required accuracy of the locations. The following table summarizes our observations:

<u>Instrument used</u>	<u>Location in</u>	<u>Reading Time in hours</u>	<u>Plotting Time in hours</u>
none	degrees	10 - 12	9 - 10
none	metres (U.T.M.)	5 - 5.5	4 - 4.5
Gerber scale	degrees	16 - 17	12 - 13
Gerber scale	metres (U.T.M.)	12 - 13	8 - 10

Accuracy

In order to determine the degree of accuracy obtained in reading the locations of 660 points from three types of maps, it was felt necessary to plot them on similar types of maps. Then the positions of the plotted points were checked against those in the original grid overlay and maps. The results are given in the following table:

<u>Instrument used</u>	<u>Location in</u>	<u>Percentage of exactly plotted points</u>	<u>Percentage of Erroneously plotted points</u>	<u>Maximum error in terms of grid spacing</u>
none	degrees	10%	90%	$\pm .15$
none	metres	20%	80%	$\pm .10$
Gerber scale	degrees	60%	40%	$\pm .08$
Gerber scale	metres	93%	7%	$\pm .03$

One takes 20 to 33 hours to read the location of 660 points in latitude and longitude from a map and then to plot the points. In the case of rectangular coordinates (U.T.M.), the required time is about 12 to 20 hours.

The present study has led us to conclude that it is much more accurate and quicker to read the location of points in terms of rectangular (U.T.M.) coordinates rather than in spherical (latitude and longitude) coordinates. Moreover, for rectangular coordinates the degree of accuracy in reading locations of points can be greatly enhanced by advantageous use of the Electronic Coordinatographs.

6. Vertical air photographs

If it is important to define a station so that it may most easily be reoccupied, reference should be made by number to the appropriate air photograph. The station may then be indicated by using a grid in millimetres centred on the fiducial cross in the centre of the photograph. Such coordinates cannot be used for other reference because of the varying scale outward from the centre of the photograph and other factors. Reoccupation of a station is therefore the only reason for digitizing a position on an air photograph.

7. The Agricultural Rehabilitation and Development Act System

This system involves a grid of 10 mile squares whose origin is southwest of all parts of Canada. Although suitable for defining areas, it would not be satisfactory for defining points. It would not be suitable for files containing data from outside Canada.

8. Local rectangular grids

For individual projects it is often convenient to lay out a local grid either on the ground or on a map for use in locating closely-spaced samples or observations. Such a grid may be incorporated into the proposed reference system by U.T.M. or geographic coordinates of its origin. However, for systematic storage of data it would probably be preferable to relocate these samples in the U.T.M. system.

This review of coordinates currently in use indicates that no one system fulfills all important requirements for digitizing geographic location of points in Canada.

Complete computer programs for converting latitude and longitude to U.T.M. and vice versa are available. If locations are determined in either system therefore, coordinates for the other may be readily determined.

APPENDIX III

ALBERTA OIL AND GAS CONSERVATION BOARD UNIQUE WELL IDENTIFIER

The following is a description and discussion of the unique well identifier adopted by the Board for use in Alberta (February, 1966).

The unique well identifier will be used at the time of licensing each well. Where a license has been cancelled the identifier will also be cancelled and may subsequently be re-assigned in the same form to a well licensed later in the same legal subdivision. In rare instances where two wells have been licensed in a legal subdivision and the second well is drilled before the first one is cancelled, the identifier of the first well will be cancelled and it will be assigned to the second well in the place of its previous identifier.

Description of the Identifier:

The identifier is built around the legal survey location which is the conventional location used in Alberta, preceded by survey system and location exception codes and followed by a zero-padding and an event sequence code. These five parts are discussed in detail in Sections A-E herein and are illustrated in the following example:

	A	B	C							D	E
Description:	Survey System	Location Exception	LSD	SEC	TWP	RGE	Direction	Meridan	Padding	Event Sequence	
Code:	1	00	1206	045	15	W	5		0	0	

In its formal form the identifier illustrated above is as follows:

100120604515W500.

If clerical handling so requires, the identifier may be recorded using visual aids as follows:

100/12-06-045-15W5/00.

In this case, however, the "slash" (/) and "dash" (-) characters must be omitted during key punching.

A. Survey System: (one character)

The Dominion Land Survey system (D.L.S. = Code 1) is used throughout Alberta. Thus, every Alberta well will be assigned Code 1.

B. Location exception: (two characters)

This code is used to uniquely identify each well drilled within a legal subdivision or quadrant of a legal subdivision, by chronological order of license dates.

(1) Spacing unit of one legal subdivision or greater

In this case, the location exception simply indicates the chronological order in which the wells were licensed for a legal subdivision.

e.g. First well drilled (by license date) is coded 00
Second well drilled (by license date) is coded 02

The majority of wells in Alberta will be coded "00" since there is usually one well within a legal subdivision.

(2) Spacing unit of less than one legal subdivision

Where wells have been drilled on a spacing unit less than 40 acres (i.e. one LSD), the first character of the location exception code specifies the quadrant of the legal subdivision in which the well is located, i.e.

<u>location</u>	<u>first character of code</u>
SE 1/4 of LSD	a
SW 1/4	b
NW 1/4	c
NE 1/4	d

The second character indicates the chronological order in which the wells are licensed, i.e.

<u>sequence and location</u>	<u>code</u>
First well drilled (by license date) in SW 1/4	b0
First well drilled (by license date) in NE 1/4	d0
Second well drilled (by license date) in SW 1/4	b2

A change in the size of the drilling spacing unit will not necessitate a change in the location exception code originally assigned.

(3) Wells drilled on road allowances

In future, wells licensed to be drilled on a road allowance will be referenced south or west of the legal subdivision indicated in the identifier. For an illustration:

<u>sequence and location</u>	<u>code</u>
The first well on the road allowance south of LSD.	S0
The second well on the road allowance south of LSD.	S2
The first well on the road allowance west of LSD.	W0

For backlog wells, the location will be revised to comply with this regulation.

(4) Wells licensed to be located in the centre of a section

A well licensed to be drilled in the centre of a section or on a boundary between two legal subdivisions, must be assigned to the nearest legal subdivision north and, if necessary, east of the well location.

C. Legal Survey location: (eleven characters)

The legal survey location refers to the location of the bottom of the well. If the actual bottom hole location is found to differ from the licensed location, the legal survey location will be corrected. The legal survey location is the conventional location that has been commonly used in Alberta in past years.

D. Padding: (one character)

Padding will be coded zero ("0") in the identifier since the Dominion Land Survey System is involved.

E. Event sequence code: (one character)

This is a one digit code used to identify sets of data that are unique in connection with certain events occurring in a well.

When a well is licensed, it will receive an event sequence code of "0" (zero). This code will identify the set of data that includes general and drilling data, the completion information for all producing horizons and the official status for the first completed zone. With the completion of other zones, or the drilling of a new portion of the hole that provides additional geologic data, event sequence codes of 2, 3, 4... 9; are issued in chronological order. Event sequence code 1 will not be used.

An event sequence code, other than zero, is created as a result of:

- (1) Deepening of a well. The licensee must apply to the Board to deepen a well to a new geologic zone. The ensuing operations may supply new geological data as well as some new basic data. This leads to a new identifier. If a well is deepened as a rework operation without penetrating a new zone, however, the event sequence code remains unchanged; the TD and new completion data, etc., will be updated in the original record.
- (2) Re-entry of an abandoned well. The licensee must formally apply to the Board to re-enter an abandoned well, to deepen, whipstock, evaluate a formation, complete a well for production or complete a well as a service well. In all these cases, new data will be obtained which is unique to the re-entry operation. Hence, the event sequence code will identify this data as distinct from previous data sets.
- (3) Whipstocked portion of a well. If a decision is made to plug back an original hole and set a whipstock to directionally drill a second

hole to a new target, then this event creates an additional set of data and a second identifier for the new hole and first completion therein. The original hole, and the first completion within the original hole if applicable, would retain the initial event sequence code. Thus, the two sets of data are separately identified. A normal directionally drilled well does not affect the original unique identifier. Coordinates and location may have to be changed, however, to correct for the bottom hole location of the whipstocked hole as previously mentioned.

- (4) Second and subsequent completions. At the time a well is officially completed in more than one pool, an event sequence code will be issued to identify the record for the completion in each pool.

All above events, 1 - 4, create new sets of data, new records and corresponding identifiers. In the case of a whipstocked, deepened portion or re-entered well record, reference may be made to the event sequence code, specifying that event, in order to obtain pertinent information. A record will contain data resulting from a particular event. (In the Board's well data system, records created by second and subsequent completions may be considered as supplemental records since they will only contain the following data: Area Code; Unique Well Identifier; Well Name; Field Name and Code; Pool Name and Code; Confidential Status; Confidential Depth; On Production Date; Dates and Status History of the well after completion. To obtain all other information, it is necessary to refer to the record coded "0".)

APPENDIX IV

RECORDED REFERENCE NUMBERING FORMAT

APPENDIX V

REPORT ON THE FREQUENCY AND DEGREE OF COMMONNESS OF THE INFORMATION RECORDED IN FIELD NOTEBOOKS

(F.P. Agterberg)

This report was prepared to meet the request of subcommittee No. 2 on Digitizing Descriptive Data of the ad hoc Committee.

The objective of the study was to estimate by first approximation the frequency and the degree of commonness of the information in sets of field notebooks of Geological Survey of Canada staff. Five areas were selected for analysis as follows:

- A: Canadian Shield; Reconnaissance study - approx. 500 stations
- B: Cordillera; Detailed study of intrusives - approx. 500 stations
- C: Sedimentary Rocks: Part of reconnaissance operation - approx. 500 stations
- D: Sedimentary Rocks: Regional study - 435 stations
- E: Sedimentary Rocks: Detailed local study - 382 stations

The notebooks were examined for the frequency with which data or other information was recorded for specific geological factors. The factors investigated were prescribed by the subcommittee, and are listed below with frequencies shown as percentage values.

	A	B	C	D	E
1. One or more samples collected from observation point	49%	14%	16%	55%	<25%
2. Geologic form specified (e.g. stock, dike)	25	8	1	1	11
3. Stratigraphic measurements in inches, feet	-	-	-	41	72
4. Information on intercalated beds	2	0	22	13	0
5. Presence of inclusions, description	18	30	5	6	0
6. Size of inclusions, inches	5	10	1	2	0
7. Inclusions as per cent of rock	4	4	0	1	0
8. Location, descriptive information	1	3	0	0	7
9. Weathering, information	6	7	73	28	16
10. Outcrops, size in feet	4	4	0	1	0
11. Colour, primary colour specified	3	2	19	9	12
12. Colour, shades specified	21	32	60	69	29
13. Texture, description (e.g. massive, sugary, granular, etc.)	11	11	16	12	5
14. Grain size, description (e.g. fine, medium, coarse)	22	24	10	61	1
15. Grain size, dimensions	1	2	0	0	0
16. Structure, reported (e.g. foliation, cross-bedding, etc.)	7	40	41	14	2
17. Structures, attitudes	10	22	0	0	0
18. Bedding, attitude	8	1	7	1	1
19. Mineralogy or mineralization, information (in addition to that in rock name, e.g. biotite-gneiss)	27	30	38	65	8
20. Rock name, specific	94.6	98	100	100	70

These details show a fair resemblance between areas A and B, and between areas C and D. This feature is explainable by the similarity in terrain and scope of study for these pairs.

The first and main difficulty encountered in this study is the variability in the unit of information. A large variety of rock types and geological phenomena may occur in approximately the same place, and thus carry the same observation number. Features such as layering, banding and inclusions can be readily described but will be difficult to store in a file system. When necessary, the descriptive items listed under the same observation number were sub-divided into a number of shorter units of information. This was done on the basis of such data as contacts between rock units.

In the case of detailed studies of sedimentary rocks, it is to be emphasized that stratigraphic height should be used as a parameter to locate layers with respect to one another in the file system.

A second major difficulty in a study of this type is that the relations between rock names and composition are not quickly assessable. The general procedure in the notebooks is to assign a name to a rock in the field, such name covering part of the mineral composition. In many cases additional information, in the form of description or data, is recorded on mineralogy.

The number of different descriptive terms (excluding names of rocks, minerals, fossils, and locations) was estimated for the areas A, B, C, and D.

The method consists of counting the number of new terms occurring in ten successive groups of 100 terms each. The results are listed below:

Successive groups of 100 terms	NEW TERMS				
	A	B	C	D	Average
1- 100	63	63	51	67	61
101- 200	52	48	26	35	40
201- 300	50	31	19	22	31
301- 400	28	32	15	29	26
401- 500	28	22	11	31	23
501- 600	26	10	20	9	16
601- 700	20	7	6	11	11
701- 800	16	5	6	9	9
801- 900	8	8	5	6	7
901-1000	21	11	12	7	13
1-1000	312	237	171	226	237

The number of new terms decreases with the number of previously checked terms. As a first approximation the number of new terms is inversely proportional to the number of previous terms, suggesting that a linear model may be useful for prediction purposes.

Reference: Becker, J. and R.M. Hayes, Information Storage and Retrieval, pp. 350-358; John Wiley and Sons, New York.

APPENDIX VI

SAMPLE DESCRIPTION CODE AS USED BY
GEOLOGICAL SURVEY OF CANADA

TABLE 1 - DESCRIPTION OF THE SAMPLE

(Use boxes 15 and 16)

<u>Box 15</u>	<u>Box 16</u>
1. Weathered <u>or</u>	1. Single specimen 2. Channel sample 3. Composite sample 4. Concentrate 5. Other
2. Unweathered <u>or</u>	1. Single specimen 2. Channel sample 3. Composite sample 4. Concentrate 5. Other
3. Hydrothermally altered <u>or</u>	1. Single specimen 2. Channel sample 3. Composite sample 4. Concentrate 5. Other
4. Laboratory treated	1. Single specimen
8. Other	
9. Unspecified	

TABLE 2 - CLASS OF MATERIAL

(Use boxes 17-18)

Boxes 17-18

- 00. Igneous rock
- 01. Igneous intrusive rock
- 02. Igneous extrusive rock
- or
- 10. Metamorphic rock
- 11. Meta-igneous rock
- 12. Metasedimentary rock
- or
- 20. Sedimentary rock
- 21. Clastic sedimentary rock
- 22. Non-clastic sedimentary rock
- or
- 30. Mineral
- 40. Samples of economic interest
- 50. Unconsolidated material
- 59. Combined stream water and sediment
- 60. Water
- 70. Extraterrestrial material
- 80. Biological Material
- 81. Plants
- 82. Invertebrates
- 83. Vertebrates
- 90. Unclassified

TABLE 3
IGNEOUS, INTRUSIVE, EXTRUSIVE ROCKS

(Use boxes 19-27)

Box 19

Geological Form

1. Batholith >40 sq. mi.
2. Stock <40 sq. mi.
3. Dyke
4. Sill
5. Flow
6. Bedded deposit
7. Volcanic fragmental
8. Vein
9. Other

Box 20

Box 21

Grain Size and Texture

- | | | |
|------------------------|---------|---------------------------------|
| 1. Glassy | | 1. Uniform grain size |
| 2. Aphanitic | | 2. Variable grain size |
| 3. Fine grained | < 1 mm | 3. Porphyritic |
| 4. Medium grained | 1- 5 mm | 4. Intersertal-ophitic-diabasic |
| 5. Coarse grained | 5-20 mm | 5. Poikilitic |
| 6. Very coarse grained | >20 mm | 6. Pegmatitic |
| | | 7. Tuffaceous |
| | | 8. Proto- or cataclastic |
| | | 9. Other |

Dominant Minerals

Boxes 22-23

Boxes 24-25

Boxes 26-27

01. Quartz
02. Plagioclase
10. Potash feldspar
20. Mica
30. Oxide-sulphide
40. Amphibole
50. Clinopyroxene
60. Orthopyroxene
70. Olivine
80. Feldspathoid
90. Other

01. Quartz
02. Plagioclase
10. Potash feldspar
20. Mica
30. Oxide-sulphide
40. Amphibole
50. Clinopyroxene
60. Orthopyroxene
70. Olivine
80. Feldspathoid
90. Other

01. Quartz
02. Plagioclase
10. Potash feldspar
20. Mica
30. Oxide-sulphide
40. Amphibole
50. Clinopyroxene
60. Orthopyroxene
70. Olivine
80. Feldspathoid
90. Other

TABLE 4

METAMORPHIC, META-IGNEOUS, METASEDIMENTARY ROCKS

(Use boxes 19-27)

Box 19

1. Contact
2. Regional
3. Dynamic
4. Metasomatic
5. Other

Grain Size and Texture

Box 20

1. Very fine-grained
2. Fine-grained
3. Medium-grained
4. Coarse grained
5. Very coarse-grained

- <1 mm
1- 5 mm
5-20 mm
>20 mm

Box 21

1. Uniform grain size
2. Variable grain size
3. Porphyroblastic
4. Poikiloblastic
5. Other
6. Foliation

Boxes 22-23

01. Quartz
02. Feldspar
10. Micas
15. Chlorites
20. Amphibole
30. Pyroxene
40. Al-silicates
50. Carbonate
60. Calc-silicates
70. Garnet
80. Other

Boxes 24-25

01. Quartz
02. Feldspar
10. Micas
15. Chlorites
20. Amphibole
30. Pyroxene
40. Al-silicates
50. Carbonate
60. Calc-silicates
70. Garnet
80. Other

Boxes 26-27

01. Quartz
02. Feldspar
10. Micas
15. Chlorites
20. Amphibole
30. Pyroxene
40. Al-silicates
50. Carbonate
60. Calc-silicates
70. Garnet
80. Other

TABLE 5

SEDIMENTARY, NON-CLASTIC, CLASTIC SEDIMENTARY ROCKS

<u>Box 19</u>	<u>Box 20</u>
Rock Type	Textural Modifiers
0. Conglomerate or breccia 1. Arenite 2. Lithic arenite 3. Feldspathic arenite 4. Wacke 5. Siltstone 6. Clay stone 7. Mudstone 8. Shale 9. Other	0. Conglomeratic 1. Arenaceous 2. Silty 3. Argillaceous 4. Oolitic-pisolitic 5. Bioclastic-skeletal framework 6. Pelletoidal 7. Stromatolitic 8. Fossiliferous 9. Sparry
<u>Boxes 21-22</u>	
Dominant Rock and Mineral Grains	
01. Mineral grains unspecified 02. Quartz 03. Feldspar 04. Clay minerals 05. Micaceous minerals 06. Phosphatic minerals 07. Chert 08. Iron oxides 09. Iron silicates 10. Rock fragments unspecified 11. Volcanic rock 12. Pelitic rock	20. Carbonate grains unspecified 21. Calcite-dolomite 22. Calcite 23. Dolomite 30. Evaporite minerals 31. Gypsum-anhydrite 32. Halite-sylvite 40. Other
<u>Box 23</u>	<u>Box 24</u>
Compositional Modifiers	Cement
0. Calcareous 1. Dolomite 2. Chert, chalcedony, opal 3. Quartzose 4. Glauconitic 5. Ferruginous 6. Gypsiferous, anhydritic 7. Phosphatic 8. Carbonaceous, bituminous 9. Micaceous	0. Calcite 1. Dolomite 2. Chert, chalcedony, opal 3. Quartz 4. Ankerite 5. Siderite 6. Barite 7. Iron oxides 8. Undifferentiated 9. None

Box 25

Mean Grain Size

0. Unspecified		
1. Clay	≤ 0.004 mm	(cryptocrystalline)
2. Silt	0.004-0.062	(microcrystalline)
3. Very fine sand	0.062-0.025	(very fine crystalline)
4. Fine sand	0.125-0.25	(fine crystalline)
5. Medium coarse sand	0.25 -0.5	(medium coarse crystalline)
6. Coarse sand	0.5 -1.0	(coarse crystalline)
7. Very coarse sand	1.0 -2.0	(very coarse crystalline)
8. Granules and pebbles	2.0 -64.0	(")
9. Cobbles and boulders	>64.0	(")

Boxes 28 - 31

Boxes 28 - 31 provide for mnemonically coded rock name, mineral name or other material.

TABLE 9
STRATIGRAPHIC AGE OF SPECIMEN

(Use boxes 32-33)

- 01. Precambrian undivided
- 02. Archaean (2400m.y.)
- 03. Archaean-Proterozoic
- 04. Proterozoic (600-2400m.y.)

- 10. Palaeozoic undivided
- 11. Proterozoic-Palaeozoic
- 12. Cambrian
- 13. Cambrian-Ordovician
- 14. Ordovician
- 15. Ordovician-Silurian
- 16. Silurian
- 17. Silurian-Devonian
- 18. Devonian
- 19. Devonian-Mississippian
- 20. Carboniferous
- 21. Mississippian
- 22. Pennsylvanian
- 23. Pennsylvanian-Permian
- 24. Permian

- 30. Mesozoic undivided
- 31. Mesozoic-Palaeozoic
- 32. Triassic
- 33. Triassic-Jurassic
- 34. Jurassic
- 35. Jurassic-Cretaceous
- 36. Cretaceous

- 40. Cenozoic undivided
- 41. Mesozoic-Cenozoic
- 42. Tertiary
- 43. Tertiary-Quaternary
- 44. Quaternary

- 50. Unknown

APPENDIX VII

CODING PROCEDURES IN NATIONAL INDEX TO GEOLOGICAL DATA

Part A: Document titles and other bibliographic data-Examples

- | | |
|---|---|
| <p>(a) <u>Title for published documents</u></p> <ol style="list-style-type: none">1. Owner's title of document2. Personal author3. Corporate author4. Date of origin5. Owner's publication number6. Availability not coded | <p>(b) <u>Title for unpublished documents</u></p> <ol style="list-style-type: none">1. Owner's title of document2. Personal author3. Corporate author4. Date of origin5. Owner's file call number6. Availability should be coded |
|---|---|
- (c) Title format for maps
1. Area name amplified by NTS number
 2. Type of map
 3. Horizon (if other than surface, and interval if applicable)
 4. Personal author
 5. Corporate author
 6. Scale
 7. Date of origin
 8. Availability
- (d) GEØDAT title format
1. Amplified title including area name plus survey procedure if significant, e.g. if project is a helicopter or geochemical project
 2. Scale e.g. 1 IN = 1 MI
 3. NTS number-e.g. nnnA, nnnA/E, nnnA/W
 4. Personal author-BY C.H. SMITH
 5. Corporate author-GSC
 6. Date of beginning-1965
 7. File identification and call number-GEØDAT 00057027
 8. Availability of file-confidential, restricted or open written in full.

Part B: Specifications of the system output control

The value of output control is to limit the volume of print out and it should be built into the system from the beginning. It is recommended that the Index should be of a type that allows for massive search and/or discrete search. It should be set up in such a way that the central operating authority would be able to provide a question answering service or supply printed out indexes either periodically or on request of the whole system or of special sections.

As a start, and realizing these are guide lines only which will be modified by experience gained in operation of the index, it is suggested that the following categories could be considered in providing output control.

(a) Province or Territory

- | | |
|-----------------------------------|-------------------------|
| 01 Newfoundland and Labrador | 08 Saskatchewan |
| 02 Nova Scotia | 09 Alberta |
| 03 Prince Edward Island | 10 British Columbia |
| 04 New Brunswick | 11 Yukon Territory |
| 05 Quebec-Anticosti, Magdalen Is. | 12 MacKenzie District |
| 06 Ontario | 13 Franklin District |
| 07 Manitoba | 14 District of Keewatin |
| | 15 Canada |

(b) Geological Eras

- 21 Precambrian
- 22 Palaeozoic
- 23 Mesozoic
- 24 Cenozoic
- 25 Other

(c) Discipline

- 41 Geophysics
- 42 Geochemistry
- 43 Geology
- 44 Palaeontology
- 45 Physiography

(d) Corporate author

- 81 Geological Survey of Canada
- 82 Ontario Department of Mines
- 83 B.C. Department of Mines and Petroleum Resources
etc.

Imperial Oil started with general guides and then through regular meetings of a liaison committee with specialist groups has modified the output controls to their present state.

In addition to the discipline categories shown above there will be a need to indicate the holdings of specific contributing groups, i.e. Ontario Department of Mines, Geological Survey of Canada, Imperial Oil, etc.

While, at first, the output control would be used to produce printed indexes, when machine search of the Index file becomes feasible it would be used to limit the amount of search to a specific discipline or the holdings of a particular group, instead of the entire file.

Part C: Type of original document and provisional group code numbers

The index should indicate the type of original document e.g.

- 01 book or textbook
- 02 published report
- 03 published map
- 04 unpublished report
- 05 unpublished map
- 06 thesis
- 07 periodic literature
- 08 magnetic tape or I.B.M. card files
- 09 assay or analysis sheet
- 10 lithologic log
- 11 field notes
- 12 memo or correspondence

Part D: Geographic location

To facilitate the use of location as a concept for index purposes it is necessary to define a specific set of areas which not only is familiar to potential users of the Index but which is capable of being modified in time to increase its specificity. For instance, in a particular defined area there may be, at the outset of the Index, half a dozen index items listed. As time goes on and more data of different kinds are gathered at different points within this area the number of index items will not only become larger but the requirements of searchers may tend to become increasingly specific. Under these conditions a search would recover a larger number of non-relevant documents.

The National Topographic System has the advantage of being well-known to geologists in general and it has a capacity of defining areas to a range of specificity that is of practical interest in the designation of index location concepts. It is recommended that location be defined as areas in terms of existing subdivisions of N.T.S. primary quadrangles. The alternate of defining areas in terms of degree "squares" has some merit but only for southern regions.

A lettered quadrangle in the N.T.S. system in southern latitudes is one degree of latitude by two degrees of longitude in size i.e. two degree "squares". Each degree "square" becomes narrower in actual width parallel to latitude as one proceeds northward and, north of 68°N, each lettered quadrangle contains four degree "squares" rather than two. North of 80°N (and there is a sizeable piece of Canada further north than this), a lettered quadrangle contains eight degree "squares". This raises complications in the labelling of degree "squares" other than by the coordinates of a particular corner and this kind of label is more cumbersome than an N.T.S. reference. The "squares" also assume a rather impractical shape in Northern latitudes.

In order to base location concepts on areas of roughly the same size and shape the basic unit of location should be a "rectangular" area consisting of half an N.T.S. lettered quadrangle; each of these should be designated as the east or west half. Thus a basic location area would be described as, for example, 63 P EAST, 560 A WEST and so on. South of latitude 68°N such a rectangle would be a degree "square".

In the future, and as the Index gets larger, such a degree square could be broken down into its component N.T.S. subdivisions but this potential requirement and its practicality need not be considered now. All the location concepts possible would be defined at the outset but the index output could be programmed to omit the printing of a location for which no data has been indexed.

For location concepts embracing an area larger than or transgressing the bounds of half of a lettered quadrangle, it will be necessary to define more than one area concept. Exception would be made for very large areas or wherever a regional name is well known (e.g. "Vancouver Island" which requires eight N.T.S. concepts to cover it). In these cases, the usual name would be used rather than N.T.S. concepts.

Part E: Preliminary guidelines for the definition of the data content of indexed documents

(a) Analyses expressed numerically

1. Rock and mineral analyses
2. Sedimentary material size analyses
3. Assays
4. Metal content

(b) Quantitative measurements on geological bodies or structural features

1. Dimensions such as length, breadth, thickness, or depth
2. Thickness of named rock formations
3. Attitudes of geological bodies and structural features

(c) Quantitative measurements of physical properties of geological materials

1. Porosity
2. Specific gravity
3. Magnetic properties
4. Seismic properties
5. Isotopic age
6. Isotopic abundances

(d) Field locations of minerals, fossils, rocks, attitudes, biological materials, type stratigraphic sections

(e) Existence of fossil name lists coupled with name of enclosing formation, stratigraphic age of formation

TABLE 1

Sample Description Format by K.R. Dawson

- (1) - (10) Specimen or observation number.
- (11) 1. Specimen, 2. Observation.
- (12) Specimen Condition: 1. fresh, 2. weathered, 3. other.
- (13) Hardness: 1. hard, 2. medium, 3. soft.
- (14) Sample type: 0. unspecified, 1. hand, 2. chip, 3. channel, 4. core, 5. cuttings.
- (15) Brilliance: 1. black to dark grey, 2. light grey to white.
- (16) Hue: 1. red, 2. brown, 3. yellow, 4. green, 5. purple, 6. blue.
- (17) General structure: 0. not recorded, 1. massive, 2. not massive.
- (18) Bedding: 0. not recorded, 1. not bedded, 2. thick beds, 3. thin beds, 4. beds of variable thickness.
- (19) - (23) Strike and dip of bedding (numerical).
- (24) Foliation: 0. not recorded, 1. not foliated, 2. simple, 3. complex.
- (25) - (29) Strike and dip of foliation (numerical).
- (30) Lineation: 0. not recorded, 1. not lineated, 2. simple, 3. complex.
- (31) - (35) Strike and plunge of lineation (numerical).
- (36) Shearing: 0. not recorded, 1. not sheared, 2. intense, 3. moderate, 4. weak.
- (37) - (41) Strike and dip of shearing (numerical).
- (42) Jointing: 0. not recorded, 1. not developed, 2. simple, 3. complex.
- (43) - (47) Strike and dip of jointing (numerical).
- (48) Folding: 0. not recorded, 1. not present, 2. simple anticlinal, 3. simple synclinal, 4. complex.
- (49) - (53) Strike and dip of axial plane of folding (numerical).
- (54) Texture: 0. not recorded, 1. not developed, 2. crystalline, 3. clastic.
- (55) Texture is crystalline: 1. porphyritic, 2. ophitic, 3. interlocking, 4. pillows, vesicles etc.
- (56) Texture is clastic: 1. angular, 2. rounded, 3. porous, 4. non-porous, 5. accretionary, 6. non-accretionary.
- (57) Maximum grain size: 1. <1 mm, 2. 1-5 mm, 3. 5-20 mm, 4. >20 mm.
- (58) Mean grain size: 1. <1 mm, 2. 1-5 mm, 3. 5-20 mm, 4. >20 mm.
- (59) Minimum grain size: 1. <1 mm, 2. 1-5 mm.
- (60) - (61) Dominant mineral (2-digit code)
- (62) - (63) Second mineral (" ")
- (64) - (65) Third mineral (" ")
- (66) - (67) Fourth mineral (" ")
- (68) - (69) Fifth mineral (" ")
- (70) Shape of rock unit is linear: 1. small < 1000' diam., 2. large > 1000' diam.
- (71) Shape of rock unit is tabular: 1. thickness unknown, 2. <10', 3. 10' - 100', 4. 100' - 1000', 5. >1000'.
- (72) Rock unit is 3-dimensional: 1. mean horizontal diameter < 6 miles, 2. mean horizontal diameter > 6 miles.
- (73) Rock classification: 1. plutonic, 2. volcanic, 3. sedimentary, 4. unknown, 5. metamorphosed.
- (74) - (77) Rock name in 4 letter mnemonic code.
- (78) - (79) Blank.

TABLE 2

Sample Description Format by N.M. Ediger

- (1) - (21) Reference number and location.
- (22) Observation = 0 or sample = S.
- (23) Surface exposure = S, subsurface = Z.
- (24) Outcrop variability: Homogeneous = U, Heterogeneous = V.
- (25) Sample: 1. represents whole exposure, 2. represents major part, 3. represents minor part.
- (26) Sample type: 1. unspecified, 2. hand specimen, 3. chip, 4. channel, 5. core, 6. drill cuttings.
- (27) Brilliance, fresh surface: 1. black to dark grey, 2. light grey to white.
- (28) Brilliance, weathered surface: 1. black to dark grey, 2. light grey to white.
- (29) Hue, fresh surface: 1. red, 2. brown, 3. yellow, 4. green, 5. purple, 6. blue.
- (30) Hue, weathered surface: 1. red, 2. brown, 3. yellow, 4. green, 5. purple, 6. blue.
- (31) Hardness, fresh surface: 1. hard, 2. medium, 3. soft, 4. loose.
- (32) Hardness, weathered surface: 1. hard, 2. medium, 3. soft, 4. loose.
- (33) Structure: 1. massive, 2. not massive, 3. not recorded.
- (34) Bedding: 1. thickly bedded, 2. thinly bedded, 3. not recorded, 4. not bedded.
- (35) - (39) Strike and dip of bedding (numerical).
- (40) Foliation: 1. simple, 2. complex, 3. not recorded, 4. see detail file.
- (41) - (45) Strike and dip of dominant foliation (numerical).
- (46) Lineation: 1. simple, 2. complex, 3. not recorded, 4. none, 5. see detailed file.
- (47) Shearing: 1. present, 2. absent, 3. not recorded.
- (48) Jointing: 1. simple, 2. complex, 3. not recorded, 4. none, 5. see detail file.
- (49) Folding observable in outcrop: 1. present, 2. none, 3. not recorded, 4. detail file.
- (50) Texture: 1. aphanitic or glassy, 2. uniform, 3. variable.
- (51) Texture, crystalline rocks: 1. porphyritic, 2. poikilitic, 3. ophitic.
- (52) Texture, crystalline rocks: 1. interlocking, 2. not interlocking.
- (53) Texture, crystalline rocks: 1. evidence of volcanic origin, 2. pillows, 3. vesicles, 4. amygdules.
- (54) Texture, clastic rocks: 1. angular, 2. subrounded, 3. rounded.
- (55) Texture, clastic rocks: 1. non-porous, 2. porous and carbonate rocks.
- (56) Texture, clastic rocks: 1. accretionary, 2. non-accretionary.
- (57) Maximum grain size: 1. < 1 mm, 2. 1-5 mm, 3. 5-20 mm, 4. > 20 mm.
- (58) Mean grain size: 1. < 1 mm, 2. 1-5 mm, 3. 5-20 mm, 4. > 20 mm.
- (59) Minimum grain size: 1. < 1 mm, 2. 1-5 mm, 3. 5-20 mm.
- (60) - (69) Dominant, second, third, fourth and fifth minerals using a 2-digit code to include the following minerals or constituents: quartz (chert), feldspar, amphibole, pyroxene, mica (sericite), chlorite (serpentine), olivine, aluminum silicates, feldspathoid, garnet, staurolite, carbonate, phosphate, glauconite, oxides, sulphides, evaporates, clays, carbon compounds, fossils, glass.
- (70) Shape of rock unit: 1. linear, 2. tabular, 3. 3-dimensional.
- (71) Diameter of linear bodies: 1. < 1000', 2. > 1000'.
- (72) Thickness of tabular bodies: 1. unknown, 2. < 10', 3. 10' - 100', 4. 100' - 1000', 5. > 1000'.
- (73) Mean horizontal dimension: 1. < 6 miles, 2. > 6 miles, 3. sedimentary, 4. unknown.
- (74) Suggested rock group: 1. plutonic, 2. volcanic, 3. sedimentary, 4. unknown.
- (75) Level of metamorphism: 1. not metamorphosed, 2. greenschist facies, 3. amphibolite facies, 4. granulite facies.
- (76) - (79) Suggested rock name - field designation, mnemonic code.
- (80) Blank.

TABLE 3

MINERAL DEPOSIT RECORD

GENERAL HEADING		DETAIL		REMARKS
IDENTIFICATION (REFERENCE NO.)				
BRIEF CHARACTERIZATION		Metal(s) major/minor Commodity (Substance, Product) Mineral(s)		Chemical symbols + mnemonic names
NAME(s)		Current Original Standing Other		MNEMONIC
LEGAL OR ADMINISTRATIVE STATUS		Claim (in good/not in good standing) Lease Other Owner(s)		((Coded (MNEMONIC
LOCATION GEOGRAPHICAL		Latitude Longitude NTS UTM		degrees to five decimals " " " " " " to 1° "square"
GEOLOGICAL		Description Mining Camp (Name) Other (isolated, particular geological unit)		verbal
ADMINISTRATIVE		Country Province or Territory Mining Division County or District Township or Parish Range Concession Lot		MNEMONIC (or coded)
ELEVATION				feet above sea level

TABLE 3 continued

GENERAL HEADING	DETAIL	REMARKS
DEGREE OF DEVELOPMENT	Former Producing Mine Producing mine Potential producer Drilled (ore reserves) Prospect (with some drilling) Prospect with trenching only	could be coded e.g. 01 to 08
HISTORY OF INVESTIGATION	Discovery date _____ by _____ total _____ Trenching date(s) _____ total _____ Drilling date(s) _____ total _____ Geological study date(s) by _____ Geochemical " " (s) " Geophysical " " (s) " Other studies	or date only or trenched/not trenched or drilled/not drilled or done or not done only " " listed mnemonically?
RESERVES GRADES PRODUCTION	Production From/to Grade Reserves Grade	four significant figs. year dates three main elements four significant figures three main elements
BIBLIOGRAPHIC REFERENCES	Published reports maps other documents Unpublished This Record prepared by _____ date _____ Revised _____ date _____ Revised _____ date _____	

REMARKS

DETAIL

GENERAL HEADING

UNIT BEING RECORDED (UNIT CONCEPT)	Simple Compound Multiple Complex Not precise	What is the "mineral deposit" being recorded
	Part of larger unit	
DOMINANTLY LINEAR	Pipe Parallel pipes Elongated lens En echelon lenses Saddle reef Other	one predominating direction i.e. two short one long $x = y < z$.
DOMINANTLY TABULAR	Vein Parallel veins. Some en echelon Dike Stratiform Parallel beds. Bedded stockworks Lens Other	
NO DOMINANT DIRECTION	Thick lens. Vein swarm. Stockworks Irregular body " " Indefinite " " Other	
ATTITUDE	Strike Dip Pitch Plunge Complex known but cannot be simply described Unknown. Insufficiently known	

TABLE 3 continued

GENERAL HEADING	DETAIL	REMARKS
SIZE		
DIMENSIONS	Unknown One known Two known Three known	
LENGTH	Horizontal (strike) Greatest (plunge)	
WIDTH	Horizontal Perpendicular to dip	
VERTICAL	True vertical Down dip	
MINERALOGY		
DOMINANT	Rock or Mineral Name (Mnemonic)	
SUBDOMINANT 1	"	
SUBDOMINANT 2	"	
SIGNIFICANT 1	"	
SIGNIFICANT 2	"	

GENERAL HEADING	DETAIL	REMARKS
WALLROCK		
COMPLEXITY		
RELATION TO DEPOSIT		
HANGING WALL (or both walls if HW & FW the same)		
	Single Homogeneous Single Anisotropic Two More than two Concordant Discordant Complex Indeterminate Complex Unknown Sediment. Clastic Sediment. Other Igneous. Unknown " Extrusive " Pyroclastic " Intrusive	
FOOTWALL		
	Complex Unknown Sediment. " Clastic " Other Igneous Unknown Extrusive Pyroclastic Intrusive	

TABLE 3 continued
DETAIL
REMARKS

GENERAL HEADING	DETAIL	REMARKS
WALL ROCK MINERALOGY		
DOMINANT		
SUBDOMINANT		
SIGNIFICANT 1		
SIGNIFICANT 2		
AGE		
MINERALIZATION		
WALL ROCKS		

TABLE 4

RECOMMENDED ENTRIES AND GENERAL SPECIFICATIONS FOR FILE ON FOSSIL FUELS AND RELATED ECONOMIC SEDIMENTARY DEPOSITS						
DATA ENTRY	APPLICABLE CLASS (A-2)					GENERAL SPECIFICATIONS
	Hydrocarbons, etc.	Coal	Evaporites	Other industrial min.	Formation water deriv.	
A. INDEX						
1. Name of deposit						
a. deposit (pool, seam or other stratigraphic unit)	X	X	X	X	X	—basic unit to which this file applies; defined, ideally, as a three-dimensional body of strata characterized by the presence of materials currently of economic value which are being, or are capable of being, economically exploited; a hydrocarbon "pool" is defined by Levorsen (1954, p. 29-30) as "the body of oil or gas or both occurring in a separate reservoir and under a single pressure system"; deposit commonly named from rock unit in which commodity occurs; file can apply to depleted deposits.
b. field, mine, or pit	X	X	X	X	X	—authorized or informal geographic name; may comprise one or more deposits.
c. regional deposit		X	X	X	X	—regionally extensive deposit, only part of which is characterized by this file; areal limits of such partial deposits arbitrarily defined, but specified.
2. Class of deposit	X	X	X	X	X	—(i) hydrocarbons and related commodities (crude oil, natural gas, helium, carbon dioxide, hydrogen sulphide, nitrogen), (ii) coal, (iii) evaporites (potash, gypsum, halite), (iv) other industrial minerals of sedimentary origin (sand, gravel, lime, clays, sodium sulphate), (v) formation water derivatives (brine).
3. Location						
a. province and/or territory	X	X	X	X	X	—10 provinces, 2 territories (up to 3 entries).
b. county and/or district	X	X	X	X	X	—legal subdivisions of (a) (up to 4 entries).
c. surveyed grid	X	X	X	X	X	—approx. centre of deposit; mod. N.T.S. (B.C.), D.L.S. (prairies), N.W.T. grid, range-concession (eastern Canada).
d. latitude - longitude	X	X	X	X	X	—approx. centre of deposit (c); degrees and 5 decimals.
e. offshore region	X	X	X	X	X	—(i) west coast, (ii) Hudson Bay, (iii) Lake Erie, (iv) Gulf of St. Lawrence, (v) east coast, (vi) other (specify); boundaries to be established.
B. HISTORICAL						
1. Discovery	X	X	X	X	X	—month, year of recognition as potential economic deposit.
2. First commercial production	X	X	X	X	X	—month, year.
3. Abandoned	X	X	X	X	X	—month, year.
4. Merged with other deposit	X	X	X	X	X	—month, year; names (A-1) of merged deposits.
5. Shut-in or suspended	X	X	X	X	X	—month, year.
6. Reactivated	X	X	X	X	X	—month, year.
7. First production stimulation	X				X	—month, year; type specified as in (D-4).
8. Reservoir used for storage	X	X			X	—month, year; stored commodity and storage volume specified.

TABLE 4 continued

DATA ENTRY	APPLICABLE CLASS (A-2)				GENERAL SPECIFICATIONS	
	Hydrocarbons, etc.	Coal	Evaporites	Other industrial min.	Formation water deriv.	
C. GEOLOGICAL						
1. Names of commodities extracted	X	X	X	X	X	—as specified in parentheses of (A-2)..
2. Mode of occurrence						
a. surface vs subsurface	X	X	X	X	X	—"surface" includes deposits mineable by surficial strip-ping. —(i) conventional pool, (ii) oil sands, (iii) oil shale. —(i) bedded, (ii) diapiric.
b. re petroleum	X					
c. re evaporites			X			
3. Physical properties of commodities						
Crude Oil (conventional)						
a. density	X					—degrees API; (i) average, (ii) maximum, (iii) unknown.
b. sulphur content	X					—(i) average % weight of S, H ₂ S, organic compounds, disulphides, (ii) unknown.
Gas (hydrocarbon or otherwise)						
c. summary gas analysis	X					—(i) average % weight of methane, ethane, propane, butane, pentanes and higher, CO ₂ , N, He, H ₂ s, (ii) not available. —(i) associated - occurs with crude oil, (ii) non-assoc. occurs alone.
d. associated vs non-assoc.	X					—B.T.U./ft. ³
e. calorific value	X					—see A.S.T.M. designation D 388-38.
Coal						—B.T.U./lb.
f. rank		X				—per cent; see MacKay, 1947, p. 9-11.
g. calorific value		X				
h. ash content		X				
i. sulphur content		X				
j. moisture content		X				
k. volatile matter		X				—see A.S.T.M. designation D 388-38; (i) Positive (strong, weak), (ii) Negative.
l. fixed carbon		X				
m. agglomeration prop.		X				
4. Estimated Reserves						
Crude Oil						
a. geologically (initial) in place	X					—(i) stock-tank barrels (1 bbl. = 42 U.S. gals. or 34.9723 Imp. gals.); from conventional pools and oil sand deposits, (ii) source, (iii) date, (iv) not released (v) unknown.
Oil Shale						
b. geologically (initial) in place	X					—(i) short tons, (ii) source, (iii) date, (iv) not released, (v) unknown.
c. grade	X					—Imp. gals./ton.
Gas						
d. geologically (raw) in place	X					—(i) MCF (thous. of ft. ³) @ std. conditions (60°F., 14.73 psia), (ii) source, (iii) date, (iv) not released, (v) unknown.
Coal						
e. geologically in place	X					—(i) short tons (2000 lb./ton), @ spec. grav. = 1.29 (McKay, 1947, p. 12-16), (ia) probable, (ib) possible, (ii) source, (iii) date, (iv) not released, (v) unknown.
Evaporites & Other Ind. Min.						
f. geologically in place		X	X			—(i) short tons of ore, (ii) source, (iii) date, (iv) not released, (v) unknown.
g. analysis		X	X			—to be specified.
Formation Water Der.						
h. geologically in place		X				—(i) stock-tank barrels, (ii) source, (iii) date, (iv) not released, (v) unknown.
i. analysis		X				—to be specified.

TABLE 4 continued

DATA ENTRY	APPLICABLE CLASS (A-2)				GENERAL SPECIFICATIONS
	Hydrocarbons, etc.	Coal	Evaporites	Other industrial min.	Formation water deriv.
5. Age					
a. Eon	X	X	X	X	X
b. Era	X	X	X	X	X
c. Period	X	X	X	X	X
d. Epoch	X	X	X	X	X
6. Name of rock unit					
a. Group	X	X	X	X	X
b. Formation	X	X	X	X	X
c. Member	X	X	X	X	X
d. Bed, "zone" or marker	X	X	X	X	X
7. Non-commodity lithologies					
a. dominant	X	X	X	X	X
b. subordinate	X	X	X	X	X
8. Fractures	X	X	X	X	X
9. Effective porosity	X				X
10. Dimensions					
a. gross thickness	X	X	X	X	X
b. net thickness of commodity	X	X	X	X	X
c. thickness of "pay"	X	X	X	X	X
d. number of discrete zones yielding commodity	X	X	X	X	X
e. area of surface trace	X	X	X	X	X
f. length max. intercept of (e)	X	X	X	X	X
g. length max. intercept 90° to (f)	X	X	X	X	X
h. azimuth of (f)	X	X	X	X	X
11. Geometry					
a. horizontal planar	X	X	X	X	X
12. Depth of burial					
a. to top of deposit	X	X	X	X	X
13. Elevation					
a. of top of deposit, approx. centre	X	X	X	X	X
b. approx. lowest	X	X	X	X	X
c. approx. highest	X	X	X	X	X
d. initial established interfaces	X				
14. Stratigraphic relations					
a. underlying strata	X	X	X	X	X
b. overlying strata	X	X	X	X	X
c. adjacent strata	X	X	X	X	X

Table 5

see attached "A suggested numeric hierarchical code for geologic time units"; note provision to express age in any of four ways.

ranks of units as defined by "Code of stratigraphic nomenclature": (A.A.P.G. Bull., v. 45, no. 5); record (d) only if commodity confined to it (them).

(i) carbonate (limestone, dolostone), (ii) sandstone, (iii) shale (incl. siltstone), (iv) other (specify), (v) unknown; list in order of approx. decreasing abundance; up to 3 subordinate lithologies.

—(i) significant open fracturing, (ii) significant filled fracturing (iii) little or none.

—per cent: (percentage porosity = pore volume/bulk volume x 100), (i) average, (ii) maximum.

—feet; total stratigraphic thickness of deposit; (i) average, (ii) maximum; qual. by (i) limits known or (ii) limits partly or entirely unknown; (iii) unknown.

—feet; cumulative stratigraphic thickness; (i) average, (ii) maximum, (iii) unknown.

—feet; cumulative stratigraphic thickness of strata yielding commodity; (i) average, (ii) maximum, (iii) unknown.

—arbitrary lower limit of thickness may be used; (i) number, (ii) lower limit (if any) used, (iii) unknown.

—(i) acres of proven area, (ii) limits defined, (iii) limits undefined, (iii) part of "regional deposit" (A-1-c).

—(i) miles.

—(i) miles.

—(i) degrees (0-180); record only if (f) exceeds (g) by 20%.

—ratio (10 f)/(10-g) (i) equidimensional (1.0-0.8), (ii) rectangular (0.8-0.2), (iii) elongate (<0.2).

—feet; measured from representative well, shaft, etc. near centre of deposit; specify well, shaft, etc.

—feet; datum sea level; measurement of same point as in (12.).

—feet; datum sea level; lowest known elevation of base.

—feet; datum sea level; highest known elevation of top.

—feet; (i) gas/oil, (ii) oil/water, (iii) transitional (i) or (ii), (iv) unknown.

(i) unit given in C-5, C-6, (ii) other unit(s), specify as in C-5, C-6 (iii) unknown; (i) conformable, (ii) unconformable, (iii) faulted, (iv) unknown.

—(i) unit specified in C-5, C-6, (ii) other unit(s), specify as in C-5, C-6, (iii) unknown; (i) interfinger, (ii) pinch-out, (iii) interfinger-pinch-out, (iv) faulted, (v) unconformity, (vi) unknown.

TABLE 4 continued

DATA ENTRY	APPLICABLE CLASS (A-2)	GENERAL SPECIFICATIONS
	Hydrocarbons, etc. Coal Evaporites Other industrial min. Formation water deriv.	
15. Structure		
a. regional strike	X X X X X	—degrees (0-180), (i) estimate of regional linear trend, (ii) variable, (iii) unknown.
b. regional dip	X X X X X	—(i) degrees-dip (0-90, normal; 90-180, over-turned), (ii) degrees-azimuth (0-360), direction of dip, (iii) variable (iv) unknown.
c. faulting	X X X X X	—(i) absent, (ii) fault(s) intercept deposit, (iii) deposit bounded by fault(s), (iv) faults probably absent, (v) faults probably present, (vi) unknown.
d. folding	X X X X X	—(i) absent, (ii) present, (iii) unknown.
16. Trap type	X	—Structural: (i) anticline, (ii) faulted, (iii) fractured, (iv) combination structural Stratigraphic: (v) clastic primary, (vi) carbonate primary, (vii) carbonate secondary (viii) unconformity below, (ix) unconformity above, (x) combination stratigraphic. Combination: (xi) combination of types, (xii) unknown (modified from Levorsen, 1954).
17. Structural Province	X X X X X	—according to forthcoming G.S.C. "Tectonic Map of Canada", e.g. Williston Basin, Peace River Arch, Michigan Basin, etc.; boundaries to be established by committee.
D. TECHNOLOGICAL		
1. Number of wells		
a. producing	X	—oil, gas, condensate, economic formation water.
b. abandoned producing	X	X
c. dry and abandoned	X	—within pool area, that penetrated producing interval.
d. shut-in	X	X
e. other	X	—injection, disposal, service, experimental, observational.
2. Well spacing		
a. dominant	X	—acre; (i) dominant spacing, (ii) variable, (iii) unknown.
3. Dominant drive types	X	—(i)solution-gas, (ii) gas-cap, (iii) water-drive, (iv) gravity, (v) combination, (vi) stimulation, (vii) unknown; list in approx. order of importance.
4. Production stimulation	X	—(i) none, (ii) water flood, (iii) thermal, (iv) gas injection, (v) steam, (vi) miscible flood, (vii) pressure maintenance, (viii) other (specify).
5. Initial (fm.) gas/oil ratio	X	—ft. ³ gas/barrel oil.
6. Mining method	X X X	—(i) shaft, room-and-pillar, (ii) longwall, (iii) surficial stripping, (iv) solution, (v) pit.

TABLE 5

A SUGGESTED NUMERIC HIERARCHICAL CODE FOR GEOLOGIC TIME UNITS

POSITION →		1	2	3	4-8
CODE NO.	↓	<u>ERAS</u>		<u>PERIODS</u>	<u>EPOCHS</u>
		1. Phanerozoic 2. Proterozoic 3. Archaean		1. Quaternary 2. Tertiary 3. Cretaceous 4. Jurassic 5. Triassic 6. Permian 7. Pennsylvanian 8. Mississippian 9. Devonian 10. Silurian 11. Ordovician 12. Cambrian	5-letter mnemonic codes (alpha) e.g. RECENT (Recent) MIOCEN (Miocene) MSTCN (Maastrichtian)
GEOLOGIC AGE CAN BE INDICATED AS:					
1. Specific age e.g. Silurian - 135					
2. Composite age e.g. Permo-Pennsylvanian - 131-132					
3. Minimum or maximum age e.g. pre-Devonian - 134 Min					
4. Limiting ages e.g. pre-Devonian, post-Cambrian - 134 Min 137 Max					
5. Unknown					

S & CS - WR 194/1

7 8 9		INDEXING DOCUMENT		1 2 3 4
		TITLE		80
02	GEOLOGY OF THE CROWSNEST COAL BASIN WITH SPECIAL REFERENCE TO THE FERNIE			2731
03	AREA, BY C. B. NEWMARCH, A. C. DEPT. OF MINES, 1953, BULL. 33			
04				
AUTHOR		30		
06	NEWMARCH, C. B.			
06				
1.0 4.3 4.4 8.3				
GROUP CODE 02 PUB. CODE 31 DATE INDEXED 40 66 MO. 07 DAY 26				
VOCABULARY CONCEPTS				
2 3 SUBJECTS 05 CROWSNEST COAL BASIN, B. C. 05 05 07 CROWSNEST COAL BASIN, B. C. COAL ANALYSES PHYSICAL PROPERTIES OF SHALE SHALE PHOSPHATE ROCK NATURAL GAS MICROSPORES CANNEL COAL KOOTENAY FERNIE B. C.		30 COAL 82A LIMESTONE ANALYSES COAL LIMESTONE PETROLEUM CORRELATION OF COAL SEAMS BUILDING STONE BLAIRMORE FM ELK FM		

Figure 2. Indexing document.

(Concepts)	(Index No.)	(Title and other bibliographic data, location of document)
NORITE	0927	TITLE, AUTHOR, ETC.
	0928	TITLE, AUTHOR, ETC.
	1107	TITLE, AUTHOR, ETC.
	1422	TITLE, AUTHOR, ETC.
	2977	TITLE, AUTHOR, ETC.
NOVACULITE	0150	TITLE, AUTHOR, ETC.
NOVA SCOTIA	0800	TITLE, AUTHOR, ETC.
	1001	TITLE, AUTHOR, ETC.
	1570	TITLE, AUTHOR, ETC. ←
	2977	TITLE, AUTHOR, ETC.
	2979	TITLE, AUTHOR, ETC.
NOWLAND J.	0015	TITLE, AUTHOR, ETC.

SOLUBILITY	0023	TITLE, AUTHOR, ETC.
	0875	TITLE, AUTHOR, ETC.
STREAM SILT SAMPLES	1106	TITLE, AUTHOR, ETC.
	0827	TITLE, AUTHOR, ETC.
	1570	TITLE, AUTHOR, ETC. ←
	3101	TITLE, AUTHOR, ETC.
	3174	TITLE, AUTHOR, ETC.
STRONTIANITE	0653	TITLE, AUTHOR, ETC.
	1895	TITLE, AUTHOR, ETC.
	4773	TITLE, AUTHOR, ETC.
STRONTIUM	0400	TITLE, AUTHOR, ETC.
	0653	TITLE, AUTHOR, ETC.

Figure 3. Format of Index showing examples of two possible pages and the result of a coordinate search. Search concepts "NOVA SCOTIA" and "STREAM SILT SAMPLES" reveal a possibly relevant document as shown.