1975/9

[redacted]

DEPARTMENT OF
MINERALS AND ENERGY

BUREAU OF MINERAL RESOURCES, GEOLOGY AND GEOPHYSICS

[redacted]

Record 1975/9

GEOLOGICAL DATA PROCESSING IN FRANCE
- THE SEMANTIC SYSTEM -

by

I.K. Kraitsowits

~~RESTRICTED~~

Record 1975/9

GEOLOGICAL DATA PROCESSING IN FRANCE
- THE SEMANTIC SYSTEM -

by

I.K. Kraitsowits

CONTENTS:

Page No:

SUMMARY

A scholarship awarded under the Australian Government Professional and Technical Exchange Scheme enabled the author to spend six months in France during 1971, with a brief visit to England, to study the procedures for the storage and documentation of geological data, primarily of subsurface samples.

The study was backed up by visits to several government, academic, and private organizations; each of these types of organization pursues its own lines of research, often specialized and divergent from those of other organizations, in which only selected data from the geological observations it makes are of immediate use to it. Consequently there was a tendency for each type of organization to adopt a piecemeal approach to solve its geological data storage and processing problems, with the result that several contrasting individual small computer systems were developed.

Realizing that much valuable geological information was being lost, or at least recorded only on files that were inaccessible to the majority of research workers, and that some geological observations undertaken by one organization were later being duplicated at great expense by another organization, the French Government, in 1959, gave powers to the Bureau de Recherches Géologiques et Minières to develop a centralized geological data storage and processing system on a national scale. Through the activities of the École Nationale Supérieure des Mines de Paris and the Centre Géologique Nationale in Fontainbleau evolved the semantic data processing system, which is based on the new concept of inverse data processing and appears to be adaptable to the storage and retrieval of the whole spectrum of geological information. As yet, little is known about the semantic system outside France, for many of the sources of information are confidential; however, the main principles of the system, and the adaptation of the system for the storage and processing of specific geological data, are presented in this Record.

As the semantic system is still in the juvenile stage at an operational level, it is too early to assess its potential as a solution to the data storage and processing problems at BMR. Instead, from the insight gained overseas of the uses and limitations of various computer systems, a review is presented of the difficulties that BMR must consider before developing an all-embracing system. Finally, recommendations are put forward for the adoption of a user-oriented machine-independent geological and geophysical total system which responds to a three-tier language that would be easily operated by both programmers and non-programmers alike, and would ensure confidentiality of some data.

1. INTRODUCTION

This Record outlines the semantic data processing system as it is used to record and process geological data on a national scale. The system is in an advanced stage of development in France where a theoretical base has been established, pilot runs completed, research projects carried out, and results obtained. An Act of Parliament ensures continuity of development and enforces the availability of all geological information. The system is now being applied to create resources inventories and to process geological data on a national scale. With the Imperial College in London, a joint development program is in operation (Project Geosemantica), and, within the framework of the International Union of Geological Sciences (IUGS) and UNESCO, international co-operation schemes have been drawn up (Codata, Cogeodata).

Both the account given concerning the semantic approach to data processing, and the opinions expressed concerning the other systems are subjective, as they are based on impressions gained during a six months scholarship in France. They represent only one side of the question, the semantic point of view. Such a restricted viewpoint was adopted for this presentation for the following reasons: (1) to show the reasons which necessitated a new approach and which led to the development of a new system; (2) to provide a base for evaluating these reasons; (3) to give an insight into the working of a system which has been specifically designed to store and process geological data and information on both national and international scales; (4) to invite criticisms - publications which are relevant to the semantic system, as it is used in France, are not yet available; and (5) to provide criticisms - computer manufacturers and service companies fully committed to the conventional fixed-format approach are eager to point out the advantages of their systems, but not the disadvantages.

As many of the sources of information of the semantic system are confidential, it has not always been possible to quote references where it would have been desirable to do so, but references to other systems which are in use in France and elsewhere, and which have been compared in the text with the semantic system, are contained in bibliographies compiled by Hruška & Burk (1971) and Burk (1973a, b).

Acknowledgements

The study trip was arranged between the Australian and French Governments through their technical exchange scheme; co-operation and assistance from BMR, the Australian Embassy in Paris, the French Embassy in Canberra, and the Centre International des Stages is acknowledged.

Without the co-operation of the personnel of the following establishments, which were visited, this report could not have been completed:

Bureau de Recherches Géologiques et Minières
- Direction du Service Géologique National, Paris
- Centre Scientifique, Orleans
- Nord-Pas de Calais Service Géologique Regional, Lille
- Aquitaine Service Géologique Regional, Bordeaux
- Pyrannées Service Géologique Regional, Pau

Sociéte Nationale des Pétroles d'Aquitaine, Pau
Centre de Recherches Pétrographiques et Géochimiques, Nancy
Institute Francais du Pétrole, Rueil - Malmaison
- Franlab, Rueil - Malmaison
- BEICIP, Rueil - Malmaison

DURIX - Service de Conservation des Gisements, Paris
Université de Paris, Faculté des Sciences
- Dept de Statistique Mathématique en Géologie

Ministère de Developpement National
- DICA (Direction des Carburrants)

École Nationale Supérieure des Mines de Paris
- Centre de Recherche, Fountainebleau
- Muséum National d'Histoire Naturelle, Paris

Institute of Geological Sciences, London
Royal College of Art, Division of Experimental Cartography
Graham Lee Publications, Geocom Bulletin, London
Royal Albert Museum, London
Imperial College, London.

2. GEOLOGY AND THE COMPUTER

The approach to computers in an organization depends on the objectives of that organization. The use of the computer should have its place in the global plan for the organization; besides data storage and retrieval, it should provide essential information for day-to-day work both now and in the future, it should treat all data and prepare composite results, it should be used to work out models usable in practice, and it should aid decision-making on both higher and lower levels.

There are two main problems with geological information: (1)geological information is generally too complex for a one-step input, and (2) because it is difficult to foresee how an organization and its needs will change with time, it is also difficult to decide which information is significant, and which should be recorded today in order to provide data for a function which might only become essential sometime in the future. A compromise is necessary.

The advantages of 'piecemeal' solutions and the development of individual problem-oriented small computer systems are immediately apparent: simplicity, suitability for a given organization, the limited amount of data which need to be treated, small initial investments, and quickly obtainable results.

The disadvantages are not so immediately apparent as they manifest themselves only at a later stage: individualistic approaches and, therefore, incompatibility, duplication of effort in overlapping interest fields, and difficulties in amalgamating the work of several sections and branches in larger projects. In the private organizations visited in France, where piecemeal solutions are used, the personnel frankly admitted that it is both difficult and expensive to gather overall data from the separately developed small systems; that it is virtually impossible to avoid duplication; that it is hard to ensure that all unglamorous tasks are performed besides the pet projects which provide immediate results; and that a return to the original records is often necessary because of a missing parameter or because the relevant data items were not considered significant previously and, were not recorded. This experience is perhaps universal: in the US, where this approach originated and is used extensively, Mr C. Pucinski of the Department of Defence estimated that in 1963, over 50 percent of the 15 millard dollars spent on research were wasted on the 'rediscovery' of already existing information.

The French argue strongly in favour of an all-embracing national system. The general governmental policy is to have all data and information available in a readily usable form, thereby enabling its maximum utilization. For example, the Centre de Recherches Petrographiques et Geochimiques in Nancy carries out 7000 complete rock analyses each year. The cost of the analyses includes the cost of obtaining the samples (field expenses), transport costs, sample preparation expenditures (wages, chemicals), upkeep of the research centre, running cost of equipment, etc. The burial of this information within an individualistic subsystem, and, therefore, the possibility of restricting this wealth only to the personnel in the Centre, would be considered contrary to the official policy in France.

The argument is even more pointed with national resources. It costs about $A 500 to obtain one foot of core from an offshore well; it is government policy neither to let that piece of core be wasted on any one analysis nor to let the results of that analysis become virtually the private property of the analyst or his section, simply because there is no system to make specific data available to the user when and where it is required. Levorsen, the father of petroleum geology, and also Dillon, the Vice-President of the Shell Corporation, believe that there are more reservoirs overlooked and lost in ill-conceived company files and systems than ever recovered.

The French aim is to reduce such losses. Several attempts were made to reach a suitable compromise between piecemeal solutions and a national 'total' system. In these attempts by the three main groups: research, private, and national organizations (BMR and equivalents) the following basic principles emerged.

Research organizations

The role of a research worker is to bring forward scientific progress, to contribute to human knowledge. To fulfil this role, he must be placed in an environment favourable for discovery and innovation where all relevant information is readily available. Information must be within his reach, he must know what is available, and he must know where and how to find it.

Given these conditions, a research worker can look for structures in the information pile, and perhaps make inferences not immediately apparent at first sight. Intuition, an open mind, and original methods must be used. In so doing, information is processed and treated.

The result of treatment will be a physical law, a geological hypothesis, a palaeogeographical map, or a model, which can be applied to reality.

Ideally, a research worker should have access to a total system, should draw his data from it, while his results should be immediately incorporated in that same total system. The research approach, therefore, favours the development of geological data processing total systems where all data are recorded: those which immediately appear to be significant, and also those which do not.

Private organizations

In contrast to research establishments, private organizations require a practical approach. For example, a petroleum geologist aims to discover the greatest number of commercial reservoirs in the shortest time, for minimum cost.

Private organizations also carry out research work, mainly to create models which reflect the relevant aspects of economic interests. Using easily and quickly available and reasonably well established data, the models may be built up to only that optimum level where an early decision can be made on the future of an exploration or exploitation project. Models may also be extended by extrapolation, interpolation, or analogy (for an early effort concerning the Coulombes oilfield in France, see World Oil, Nov. 1963).

Projects may be flexibly planned in advance of their operation, taking into consideration environmental restrictions (economic and political), as well as available resources (human, material, and financial).

For these types of economically oriented research work, private enterprise prefers piecemeal computer systems. All the private companies visited use conventional fixed-format systems to provide management with material for decision-making. These systems are by definition well suited for the job at hand (each is individually designed), though they are often incompatible with each other, even within the same organization.

National organization

In France, a national organization has four roles:

(1) to be a custodian and administrator of all data and information which are available in France

(2) to make its data storage, treatment, and retrieval system easily accessible to the public, to research establishments, and to private organizations. To further the development of science and exploration by co-operation and leadership

(3) to create models and to carry out applied research projects useful for further research, exploration, and administration

(4) to further international relations concerning geology. This is especially important in France, where geological basins extend into other countries, and who is a member of a union of nations (the European Common Market) and host to such international organizations as UNESCO, IUGS, etc.

These roles favour the development of a national data processing system, rather than several piecemeal solutions to restricted problems.

Developing a national system

On the 23.10.1959, the French Parliament passed an Act, which gave powers to the Bureau de Recherches Geologiques et Minieres (BRGM) to establish a geological data storage and processing system on a national scale. All geological data and information should be fed into this system, and results drawn from it. There are three problems.

The first problem is the centralization of all information. The different systems of the various governmental, semi-governmental, and research organizations should be included in a single, all-embracing system in which all subsystems are not only compatible with one other but also complement one other. The proposed scheme is shown in Figure 1. France is well on its way to achieving this aim: the framework has been established, and the system is already in operation. The co-operation of industry and the general public is ensured by the 'Code Minier' (Common Mining Code), which requires all subsurface works to be declared and all geological information to be submitted to BRGM.

The second problem is retrieval. The system should provide administration with statistical data, research with basic information, the government with resources inventories, industry with production figures, and the general public with all the released data. These are achieved through interlinked computers, and by the automatically produced 'technical files'. The technical files are both the means and the by-product of BRGM computerized semantic input system (see Fig. 1), which are released to the public to aid further exploration.

The third problem is computerization. Geological information being so variable in type, quality and quantity, there was considerable doubt in France whether the task could be performed by computers. The following advantages and disadvantages had to be considered.

Advantages of using a computer

(1) A computer can rigorously apply (and therefore amplify) that human intelligence and effort which has been fed into it.

(2) Processing times are reduced. A computer can carry out operations and calculations much faster, and more accurately, than any other method.

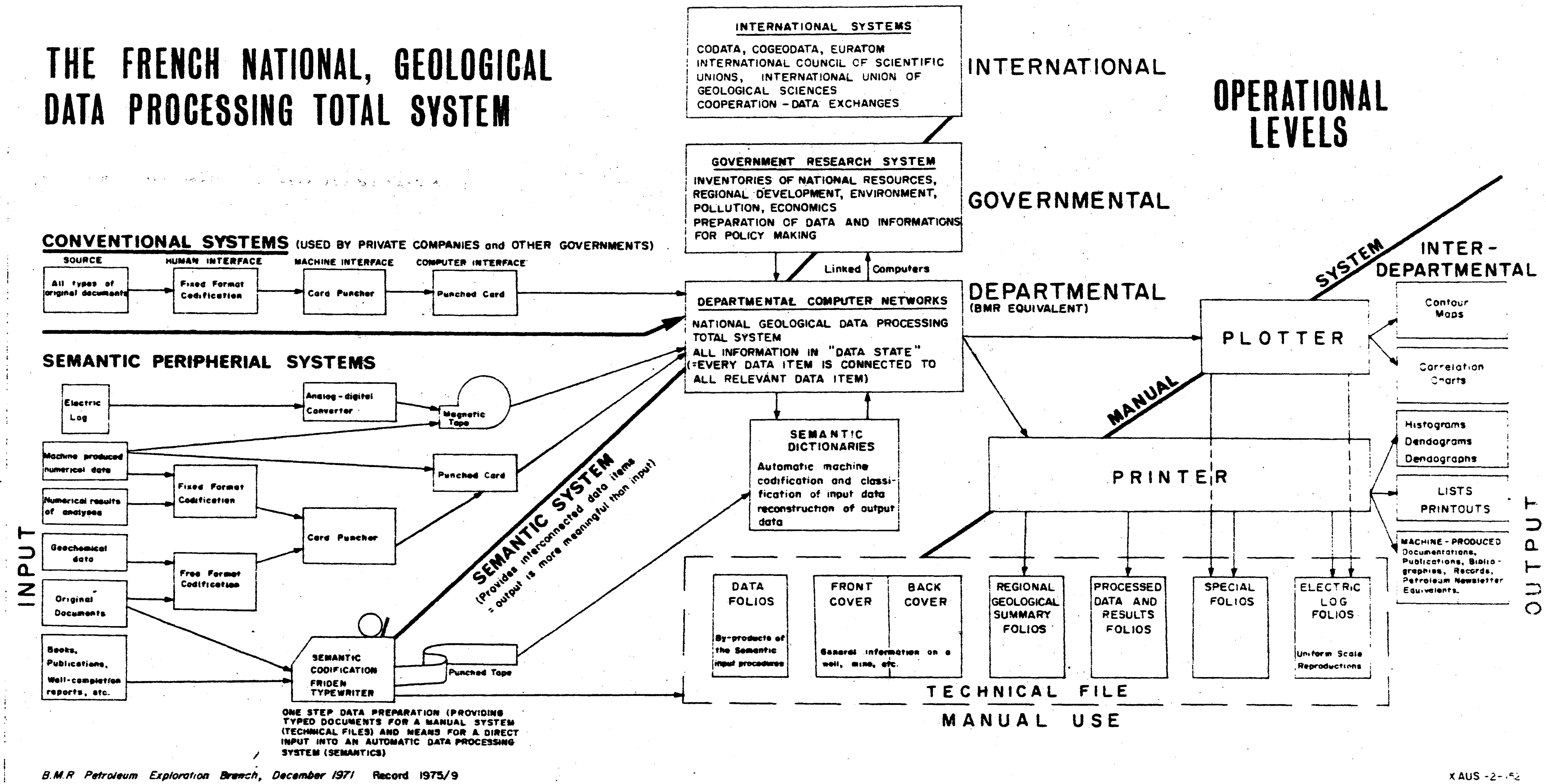

THE FRENCH NATIONAL, GEOLOGICAL DATA PROCESSING TOTAL SYSTEM
OPERATIONAL LEVELS
INTERNATIONAL SYSTEMS
CODATA, COGEODATA, EURATOM
INTERNATIONAL COUNCIL OF SCIENTIFIC UNIONS, INTERNATIONAL UNION OF GEOLOGICAL SCIENCES
COOPERATION - DATA EXCHANGES
INTERNATIONAL
GOVERNMENT RESEARCH SYSTEM
INVENTORIES OF NATIONAL RESOURCES, REGIONAL DEVELOPMENT, ENVIRONMENT, POLLUTION, ECONOMICS
PREPARATION OF DATA AND INFORMATIONS FOR POLICY MAKING
GOVERNMENTAL
Linked Computers
DEPARTMENTAL
(BMR EQUIVALENT)
INTER-DEPARTMENTAL
SYSTEM
MANUAL
CONVENTIONAL SYSTEMS (USED BY PRIVATE COMPANIES and OTHER GOVERNMENTS)
SOURCE
HUMAN INTERFACE
MACHINE INTERFACE
COMPUTER INTERFACE
All types of original documents
Fixed Format Codification
Card Puncher
Punched Card
DEPARTMENTAL COMPUTER NETWORKS
NATIONAL GEOLOGICAL DATA PROCESSING TOTAL SYSTEM
ALL INFORMATION IN "DATA STATE"
(=EVERY DATA ITEM IS CONNECTED TO ALL RELEVANT DATA ITEM)
SEMANTIC PERIPHERIAL SYSTEMS
INPUT
Electric Log
Analog-digital Converter
Magnetic Tape
Machine produced numerical data
Punched Card
Numerical results of analyses
Fixed Format Codification
Card Puncher
Geochemical data
Free Format Codification
Original Documents
Books, Publications, Well-completion reports, etc.
SEMANTIC CODIFICATION FRIDEN TYPEWRITER
Punched Tape
ONE STEP DATA PREPARATION (PROVIDING TYPED DOCUMENTS FOR A MANUAL SYSTEM (TECHNICAL FILES) AND MEANS FOR A DIRECT INPUT INTO AN AUTOMATIC DATA PROCESSING SYSTEM (SEMANTICS)
SEMANTIC SYSTEM
(Provides interconnected data items = output is more meaningful than input)
SEMANTIC DICTIONARIES
Automatic machine codification and classification of input data reconstruction of output data
PLOTTER
PRINTER
OUTPUT
Contour Maps
Correlation Charts
Histograms
Dendograms
Dendographs
LISTS
PRINTOUTS
MACHINE-PRODUCED Documentations, Publications, Bibliographies, Records, Petroleum Newsletter Equivalents.
DATA FOLIOS
By-products of the Semantic input procedures
FRONT COVER
BACK COVER
General information on a well, mine, etc.
REGIONAL GEOLOGICAL SUMMARY FOLIOS
PROCESSED DATA AND RESULTS FOLIOS
SPECIAL FOLIOS
ELECTRIC LOG FOLIOS
Uniform Scale Reproductions
TECHNICAL FILE
MANUAL USE
B.M.R Petroleum Exploration Branch, December 1971 Record 1975/9
X AUS-2-152

Fig. 1

(3) Storage space requirements are minimized. At the Centre de Recherches Pétrologiques et Géochimiques in Nancy, for example, the equivalent of two million printed pages of geochemical data are stored in one memory cell. Each cell occupies about two cubic feet of space.

(4) The information stored in a memory cell does not deteriorate with time. From the archival point of view, therefore, this method of storing data provides more accuracy than any other method.

(5) Computers diminish distances, which makes decentralization possible. Data may be transmitted through a cable, telephone, or radio. Seismic surveys digitized in the Western Australian desert can be processed and interpreted on the same day in Paris. Schlumberger Inc. intends to process data in real-time by teletransmission via satellite.

Disadvantages of using a computer

(1) The computer lacks imagination. This in turn has three consequences:

(a) the machine has to be made to test its own operation constantly. This requires involved programming.

(b) It has to be provided with data and with a program to treat the data. Data input and programming are governed by very strict rules. Meaningful results cannot be obtained until all errors have been corrected.

(c) Information has to be transcribed into a 'metalanguage', which is then deformed for input into a machine readable form. This is time-consuming and expensive. Furthermore, information may be lost owing to subjectivity.

(2) Computers are often misused, especially when a piecemeal approach to data processing is employed: the machine duplicates processes and reworks data for different users, thus diminishing the machine potential.

(3) Automatic geological data management is particularly difficult. Geological information is diverse, complex, and abundant. Terms may be poorly defined, misused, or have different meanings in different contexts. A large part of the science is descriptive, and cannot be easily formalized. Models are difficult to make and test. The basic geological reasoning is complex, and the computer cannot use intuition.

(4) Computers tend not to operate very satisfactorily. They tend to go down in the middle of jobs. The longer the job, the more likely that the computer will go down before completion of the job. Down time is due to a variety of things with a large system; power failure, disc/ tape fail, operator error, 'bad' data, power fluctuations etc.

Solutions

Many organizations adopted a piecemeal approach to data processing: different methods were used for different problems. This way, geochemical data can be stored and classified by statistical analysis methods; seismic traces can be digitized, electrical 'noises' removed by filtering, and three-dimensional models can be built using 'holoseismics'; graphic lines can be traced (maps, charts, logs, etc.) and a plotter can be used to draw contours, cross-sections, and block diagrams; factional, discriminant, cluster, and multivariance analysis techniques can be used in hydrogeology, palaeontology and stratigraphy; simulation models can create cross-sections, establish the boundaries of a palaeogeographical environment, or give pointers to the behaviour of a reservoir under actual or hypothetical conditions.

Each of these methods is different: they use different approaches, formats and standards, and they all have different input requirements. As the aims are also different, each operates only within a relatively restricted range. Therefore, none of them is immediately suited to be extended to form a base of a uniform, centralized 'total' system which can process all geological data.

In addition, none of these methods is designed for or capable of recording the descriptive information which is so common in geology. Because they all employ fixed-format input techniques (to facilitate the use of punched cards and to conform to hardware requirements), they are ill-equipped to deal with geological expressions. The coding of the geological language for computer input would require a much less formal and a more flexible approach than the one initiated by Herman Hollerith for numerical data over eighty years ago.

3. SOURCE DATA AUTOMATION

PRINCIPLES

(1) Computer-readable data should be prepared by the person who works with the data and creates the first written document. He is better qualified to judge the content and the accuracy of the data than 'data-translators', key-punchers, and key-verifiers.

(2) The codification of source data requires the transformation of all information into data items, and the coding of all non-numerical data. When original information needs to be continuously transcribed this way by someone, the result is duplication, not automation.

(3) In geology, texts (descriptions) are used almost as much as data (measured values). When a text is converted into values by an indexing process, a certain degree of uncertainty - and also subjectivity - is always present.

(4) The machine processing of descriptive information requires language formalization. Before a language can be formalized, however, the various information, data, and language (semantic, syntactical) elements need to be identified and determined.

The nature of geological information

In general, information describes a property (or characteristic) of either an object or a concept (term). Geological 'objects' include basins, formations, deposits, orebodies, deltas, mountain ranges, reservoirs, fossils, rocks, etc. Concepts include folding, metamorphism, transgression, precipitation, etc. A property might refer to the object itself (intrinsic characteristics), or to its surrounding (extrinsic characteristics). Each property may be further subdivided into: (1) an identification (the number of a rock sample or the depth of a core); (2) a name (of the characteristic described, or an identification for the value stated); and (3) a value (for the characteristic described).

Example: In sample No. 129 (identification), the percentage of silicates (name of the characteristic) is 73.79 percent (value). The sentence is one information, which contains three components (three intrinsic characteristics).

The nature of geological data

A data item refers to one observation, property, or characteristic. Dip measurements, fossil determinations, values referring to silicate, etc., percentages are data items.

A single measurement or property on its own is distorted and incomplete: one measurement cannot represent the whole range of values at a locality. A single locality could perhaps be accurately described by a large variety of data, but, in practice, this variety is reduced by selection: 'check lists' are drawn up, and observationa are 'normalized'. The range of data is further reduced by the limitations of the equipment used to express properties as numerical values. Only parts of the original complex information can, therefore, be represented as identifiable or measurable information components (data items). These can be expressed as numerals, codes, or key-words.

Manipulation of geological data and information

In some single-purpose and relatively basic and simple geological data processing systems, the relation between the various data items (objects, characteristics, and values) of an information can be expressed in the following three forms:

Logical matrices: a character is either present or absent. The character values are bivalent. The presence or absence of a character is expressed as 1 or 0.

Character Name / Object identification	A	B	C	D	E	etc
No. 1	0	0	1	1	0	...
No. 2	0	1	0	0	1	...
No. 3	0	0	1	1	0	...
No. 4	1	0	0	0	0	...
etc	...	...	...	...	...	

Fig. 2

A = presence of andalusite

B = presence of kyanite

C = presence of sillimanite

D = presence of cordierite

E = presence of staurolite

<u>Scalar matrices</u>: the characters are quantitative, multivalent, or continuous. The character values are expressed as numerals.

Character / Object Name identification	SiO_2	Al_2O_3	Fe_2O_3 Total	FeO	etc
No. 1	71.25	20.33	01.29		...
No. 2	68.74	15.42	05.85	02.10	...
No. 3	73.80	21.22	00.92	00.02	...
etc	...	...	...	...	

Fig. 3

<u>Semantic matrices</u>: the characters are geological concepts, expressed as words or sentences. Each term used in geology is a 'key-word' in semantics. The values of the words are determined by their use in the geological language.

Character / Object Name identification	Colour	Texture	etc
No. 1	Greenish	Ophitic	...
No. 2	Beige	Saccharoidal	...
No. 3	Mesocratic	Diabasic	...
No. 4	Yellow	Oolitic? Pisolitic?	...
etc	...	...	

Fig. 4

In using these matrices, preparation of information consists of placing a data item into a matrix. Uncertainty may be expressed by the establishment of numerical or semantic threshold values (CaO is between 10% and 20%;

stratigraphic age is between Permian and Cretaceous; etc); or by a question mark in the semantic matrix (sodalite?). In the matrix, various shades of uncertainty can be expressed: for example, a question mark alone might signify that a character has been observed (matrix: present), but its identity and value could not be determined.

However, not all geological observations or measurements can be expressed in these simple matrices. Often there is no clear-cut 'pigeon-hole' for a data item, or in-sufficient information is available to fill in a data item. For the input of descriptive information, further determinations are required. The geological language needs to be 'modulated' and 'formalized' before data processing on a larger scale could be attempted.

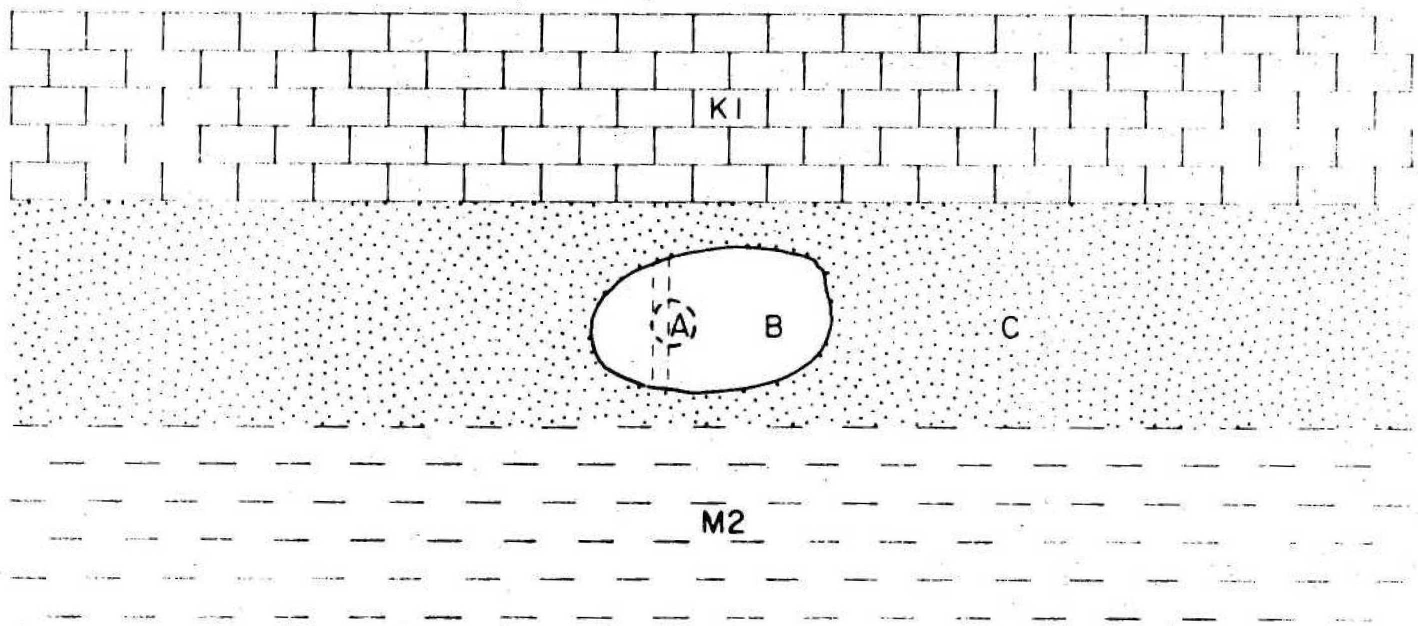
The modulation of geological information

Naming formations is straightforward if the formations in an area under consideration are well studied and named consistently, either in a formal or in an informal fashion. This is seldom so in basin studies in Australia, however, where the various source documents (company reports) use different nomenclatures.

For example, in building up a complete stratigraphy of a basin with normal geological methods, intuition plays a great part. From two sources, one combination (thought to be the most significant) is tried. If it fits, the search has ended for this particular goal of establishing a singular link between two references or two wells. If it does not fit, others are tried. From here onwards, the work proceeds according to a hypothetico-deductive method, working from the top downwards by dividing and further dividing. The procedure produces quick subdivisions at the price of several processings. The subdivisions are recognized on a progressively smaller scale. However, the method breaks down at the tail-end, where the trial and error (guess, try, and guess again) method proves to be prohibitive owing to the increasing number of combination possibilities.

On the other hand, dichotomic or 'pyramid' types of data processing can be carried out the other way round, proceeding from the bottom upwards. According to the Canadian method (Robinson, 1966), the following procedure is followed:

> Each sample examined by a sedimentary petrologist is viewed as a module within a hierarchical, pyramid-type of structural framework. (see Fig. 5). The extrinsic relation between the sample and its environments (local and regional stratigraphy) are expressed by one numbering system, which is a combination of alphabetical and



Record 1975/9

M(P1)167

Fig 5

numerical characters. The other numbering system refers to the intrinsic characteristics - the interrelations between the components within a module. The intrinsic numbering system recommences for each sample (module), with lower numbers given to small units and higher numbers to larger units. For example (see Fig. 5), in a thin-section description, if a lithic fragment (A) is found to be within a pebble (B) which is embedded in a sandstone band (C), (A) might be given the number 2, while (B) is made equal to 8 and (C) to 13. The extrinsic, or 'extra-modular' relations are coded during a macroexamination of the whole core from which the thin section was taken. If the sand-stone band is sandwiched between a limestone above and a shale below, the former might be allocated with a code K1, while the latter M2.

The object of this kind of codification is to break down the strati-graphy into a hierarchy of modules. The advantage of doing this is that it is simpler to oversee and mentally embrace a module than the whole strati-graphy. Each module is easier to code and to test, and this advantage outweighs any other.

With such a modular approach the computer processing is also simpler. Once all the components of a module (which may or may not constitute a true formation at a later stage) are identified with such a sequential numerical, alphabetical, or alphanumerical code, a sorting program can effect a sequence. The first number (or group of numbers as in the above example) are compared by the computer with the second; if the first is of a higher order, the numbers are exchanged, but not otherwise. Then the second number is compared with the third, and so on until the second last is compared to the last. The resulting stratigraphical sequence is more detailed and more accurate than that of the hypothetical-deductive method, and it can provide much more complete infor-mation concerning successions and breaks in the sequence, whereas it is not confused by local unconformities and nondepositions at the various locations.

The formalization of geological information

The quality of information preserved in a data processing system depends on the degree to which a few selected information parts (modules) can express a concept or an original text. The modules need to be determined; texts have to be analysed and summarized in such a way that the maximum information is conveyed with minimum words. These words then need to be indexed: formalized in such a way, that a machine can understand them. Obviously, not all words (concepts) can be preserved, only those which can be formalized.

The formalization of information - and, therefore, its transformation into data - is not an essential part of all automatic information systems, but in many of them, especially where indexing and codification is employed, it is the most important one. In these systems an index or a code links two essential components: information on the one hand, and its place in a classification structure on the other. A rock name, or its index or code, does not provide a complete description of everything known about that rock; in those systems which are built in conjunction with a classification structure, however, the index can represent a summary of all principal properties, if the numbered position of an item in a group, as well as the number of the group in a classification scheme, is included, or referred to in an index.

Machines can process letters and numerals, and, therefore, can be made to understand indices. This facility, besides the ability to find and identify an item by its place in a classification structure, can be used to document, store, process, and retrieve information.

<u>Example</u>: The Institute of Geological Sciences (IGS), in the United Kingdom, is using a petrological/mineralogical code, which is part of a fixed-format hierarchical classification system. IGS Report No. 70/6 gives the complete code list of most rock names used in the United Kingdom, as well as some economically or internationally important rock names (see Fig. 6); it omits locally and colloquially used names.

The rocks are arranged in a hierarchical structure, and each name is fixed-field codified. The code consists of four characters, of which the first character is the lowest level of information, referring to the following major divisions:

0: undifferentiated

1: igneous rocks

2: metamorphic rocks

3: sedimentary (clastic) rocks

4: sedimentary (chemical-biochemical) rocks

5: soils

6: mixed mineral assemblages

7: minerals

8: extra-terrestrial materials

II. ALPHABETICAL THESAURUS

IGNEOUS ROCKS
(1000)

Absarokite	USE Trachybasalt	1EG0
Adamellite		1AAB
Agglomerate, undiff., or with mixed lithic constituents		1JA0
, Amphibolite		1JAK
, Andesite		1JA5
, -porphyrite		1JA5
, Basalt		1JA9
, , undersaturated		1JAB
, Breccia		1JAN
, Carbonatite		1JAE
, Conglomerate		1JAN
, Dacite		1JA4
, -porphyry		1JA4
, Diorite		1JA2
, Dolerite		1JA9
, , undersaturated		1JAB
, Dolomite		1JAQ
, Gabbro		1JA8
, , undersaturated		1JA8
, Gneiss		1JAJ
, Granite		1JA2
, Granulite		1JAI
, Hornfels		1JAL
, Igneous constituent, undifferentiated or mixed		1JA1
, Lamprophyre		1JAD
; Limestone		1JAQ
, Metamorphic constituent, undifferentiated or mixed		1JAH
, Mineral constituent; mixed mineral and rock		1JAR
, Mineralized rock		1JAR
, Mudstone		1JAP
, Pyroclastic rock, undifferentiated		1JAG
, Rhyolite		1JA3
, -porphyry		1JA4
, Sandstone		1JAN
, Schist		1JAL
, Sedimentary constituent, undifferentiated or mixed		1JAM
, Shale		1JAP
, Siltstone		1JAP
, Slate		1JAI
, Spilite		1JAA
, Syenite		1JA2
, Syenogabbro		1JA8
, Trachyte		1JA6
, , undersaturated		1JA7
, - porphyry		1JA6
, , , undersaturated		1JA7
, Trachybasalt		1JA9
, Ultrabasic rock		1JAC
, Vitric constituent		1JAF
Agpaite	USE Undersaturated syenite undifferentiated	1CC0
Ailsyte	USE Sodic amphibole-/pyroxene-microgranite	1ABE
Åkerite	USE Melasyenite	1CAE
Alaskite		1AAN
Albitite	UF Albitophyre	1CAC
Albitophyre	USE Albitite	1CAC
Allivalite		1EAH
Alnöite		1GCE
Alsbachite	USE Microgranodiorite	1ABC

Fig. 6

SEDIMENTARY (CHEMICAL-BIOCHEMICAL) ROCKS

COL. (i): 4

COL. (ii)		COL. (iii)	COL. (iv)
N	Fe-Sulphide Rocks	0 Undifferentiated 1 Pyritic rock 2 Marcasite 3 Pyrite/Marcasite 4 Fe-sulphide-Fe-oxide/hydroxide association 5 " " Fe-silicate association 6 " " carbonate association 7 " " rudite/arenite association 8 " " argillite association;Ironstone	0 Undifferentiated 1 Rudaceous/arenaceous lithology 2 Argillaceous lithology 3 Argillaceous/arenaceous lithology 4 Silica/silicate association 5 Carbonate association 6 Phosphate association 7 Sulphate association 8 Organic compound association 9 Concretion / nodule / spherulite
P-R : Carbonaceous, Bituminous rocks and compounds — P	Carbonaceous Rocks;Coal Series	0 Undifferentiated 1 Lignite 2 Brown coal 2 Sub-bituminous coal 4 Bituminous coal 5 Semi-bituminous coal 6 Semi-anthracite 7 Anthracite 8 Humic coal 9 Sapropelic coal A Cannel coal B Boghead coal;Torbanite C Jet D Coal-igneous rock association E Coal-rudite/arenite association F Coal-argillite/arenite association G Coal-carbonate rock association	0 Undifferentiated 1 Rudaceous/arenaceous lithology 2 Argillaceous lithology 3 Calcareous lithology:Calcite 4 Carbonates:Dolomite 5 Carbonates:Ankerite;Siderite 6 Phosphate 7 Sulphate:Jarosite 8 Sulphide:Pyrite 9 Oxide/hydroxide of Fe A Durain B Fusain C Micrinite D Vitrain E Plant components (undifferentiated)
Q	Bituminous Compounds, Rocks:Solid/Plastic	0 Undifferentiated 1 Pyrobitumen 2 Asphaltite;Asphalt 3 Bitumen 4 Elaterite 5 Ozokerite 6 Hatchettite 7 Copalite 8 Bituminous compound-rudite/arenite association 9 Bituminous compound-carbonate association A Bituminous compound and ore-minerals	0 Undifferentiated 1 Rudaceous/arenaceous lithology 2 Argillaceous lithology 3 Silica/silicate 4 Silica/silicate:Clay minerals (undifferentiated) 5 Carbonate 6 Phosphate 7 Sulphate 8 Sulphide 9 Oxide/hydroxide A Organic compound (undifferentiated)
R	Bituminous Compounds:Liquid	0 Undifferentiated 1 Petroleum Series (undifferentiated) 2 " " Normal Paraffin Series 3 " " Isoparaffin Series 4 " " Naphthene Series 5 " " Aromatic Series 6 Sapropels 7 Petroleum-rudite/arenite association 8 " argillite association 9 " carbonate rock association	0 Undifferentiated

For each of the divisions shown by the first character, the second, third, and fourth characters provide successively more and more detailed information. The subdivisions in these columns are expressed either by a numeral or by a letter depending on the number of subdivisions required in each category:

2KPK = migmatitic pelitic gneiss with abundant sillimanite and some microcline

2QDO = metamorphic rock, with only one recognizable mineral: garnet

5004 = soil, categorized only from its engineering properties as fine-grained, cohesive, clayey, and of medium plasticity.

Figure 7 shows part of the sedimentary rock hierarchical structure. As the code is designed for automatic data handling, the column numbers in the display refer to those column numbers of an 80-column punched card where the code characters should be punched.

METHODS

Taxonomy is a statistical method for the systematic classification of information. Taxonomic methods are used to build documentation systems. The various existing documentation systems are designed to (1) respond to routine (frequent and identical) enquiries (library systems), or (2) satisfy research workers by answering occasional, but pointed queries (research-type retrieval systems). The former requires the documentation of types which can be classified; the latter needs to give individual attention to each input item. The former is based on the principle of selective dissemination, while the latter is based on retrospective search. In any one system, the account can be put on either one of these principles, but not on both, as each requires a different input approach.

A national system would require selective dissemination, as well as retrospective search facilities. For a single system to provide both, not only each information but each information component would have to be treated individually at the time of input and after. Existing taxonomic systems are not equipped to do this. Their advantages and disadvantages, from the point of view of their adaptability to process geological data on a national scale, are outlined below. The semantic system, which is designed to treat information as well as each information component is described in Section 4 (semantic data processing).

(1) Selective dissemination of information

Hierarchical classification systems, keyword systems, and the decimal grid system belong to this category. In these systems, both the building-up of a data bank and retrievals are based on the principle of selective dissemination of information. They are primarily designed as documentation systems - they can furnish each user with a list of references which is relevant to his field of interest. The various systems in this category differ in their method of treatment (based on coded symbols or keywords), in amount of detail (coverage), and in the degree in which interralated subjects are connected (relevance). In the hierarchical classification systems the accent is on the treatment facility, which is provided by coded symbols; in the keyword systems, on coverage; and in the decimal grid system, on relevance.

(a) Hierarchical classification systems. The various hierarchical classification systems are listed in the Appendix 1. In all of them, the selective dissemination of informations into classifiable categories is possible when the structure is organized on three or four levels as shown below.

Example: The hierarchical structure of a subject (the sea-floor) and its numerical codification -

0. The physiography of the sea and sea-shore

01. Submarine relief

011. Continental margin relief

0111. Coastal relief

0111-1.Coastal channels

0112. Continental slope relief

0112-1. Submarine canyons

012. Abyssal relief

0121. Abyssal plains

0122. Abyssal trenches

013. Epicontinental relief

0131. Submarine mountain chains

02. Littoral relief

.... etc.

Retrospective search facilities in hierarchical classification systems would require at least ten levels however. The construction of a hierarchy for all subjects on ten or more levels would make the process of codification and input prohibitively slow and difficult.

(b) Keywords. For creating bibliographies and documenting publications in a library, natural words and names, not numerals or coded symbols, need to be processed. Keyword systems fulfil this need, as in these systems words (index terms) are recognized by the computer. Only a limited number of words can be used; the size of the vocabulary depends on (1) the chosen method of formalization (fixed or variable length input formats), (2) length and number of words used (mnemonic, keyword, and natural word systems), and (3) the available memory space.

In keyword systems the information concerning the subject(s) of a publication is disseminated by using a few keywords. The same keywords as retrieval handles will produce a reference to that publication.

The index terms in keyword systems are generalized words, with a 'global' meaning. Using a limited number of these words, a thesaurus is created in which all the usable terms are arranged in a hierarchical classification structure. The thesaurus is encoded by a special Cobal program into a table which forms a computer dictionary. When details of a document and the relevant list of index terms enters the system, each term is coupled with the physical address of the document and with the code of all index terms concerning the same document. Using any one or a combination of index terms for a selection, the document number, author(s), title, date, list of index terms, citations, and a complete abstract may be retrieved. In some systems only a few index terms can be used to describe a document, but in advanced systems (for example in GYPSY = Generalized Information Processing System) up to 21 fields and an unlimited number of conceptually different index terms of undetermined (variable) length may be used.

In the Keyword - In - Context (KWIC) Library indexing systems the keywords in the title are processed and alphabetically arranged in different combinations to provide reference lists. KWIC is designed for a fast visual scan to convey the contents of the documents being indexed. The output consists of the index proper, bibliography, and author index.

In keyword systems each input/output item is immediately intelligible. This is advantageous on input, as it facilitates the dissemination of information. At the same time it is also disadvantageous, however, as it provides a temptation to by-pass the thesaurus. The inconsistent or unco-ordinated use of keywords can lead to unsuccessful retrieval attempts. The proper use of a well thought-out, and hierarchically arranged thesaurus can overcome this problem.

The principal impediment is the 'one problem, one system' approach. In palaeontology, for example, a documentation system might be based on a bibliogical classification system, whereas in other disciplines different classification systems are used. Some are structured 'from above' (palaeontology), whereas others are structured 'from below' (Sokal & Sneath, 1963). Because of this, and owing to the interdisciplinary nature of most branches of geology, terminology causes continuous problems: therefore, keyword systems based on hierarchical classifications can become out of date as soon as they are implemented. The lack of literary search facilities are also disadvantageous in keyword systems. In these systems, each keyword is a 'global' term by necessity, with a generalized meaning. 'No matter how many relations are built into this type of system, it cannot include every term and every possible relationship for every word' (Vickery, 1965). Keyword systems, therefore, can be used to disseminate information, but they are not suited for retrospective search concerning pointed issues.

Keyword systems are developed primarily for the documentation of publications and for finding references concerning a certain subject. In these systems, the maximum coverage of the subjects are more important than relevance: the accidental retrieval of marginally relevant references are tolerated for the sake of completion. Such a method would not satisfy a mineralogist or a geochemist; therefore, keyword documentation systems are also unsuitable for processing all geological data on a national scale.

A keyword system is used in France by the Centre National de la Recherche Scientifique (CNRS), which publishes the Bulletin Signaletiques, a monthly abstracting service. Three issues cover all publications in the three main fields of geology: (1) mineralogy, geochemistry, extra-terrestrial materials, and petrography, (2) applied geology, and (3) general geology and palaeontology. In this system, a Siemens 4004-35 computer (located at the Imprimerie Nationale) and a Digiset photocompositor are used. An output example is shown in Figure 8.

(c) The decimal grid. This method was originated by Messrs Van Dijk and Van Slipe, adopted by EURATOM, and developed into a complete bibliographical documentation system by BRGM. In the system, semantically related keywords are organized into hierarchical structures and graphically displayed within the framework of a decimal grid; hence the name.

N Bulletin signalétique　　Auteur　　Affiliation

34-101-699. *BUFFET (P.).* (Cent. Doc. C.N.R.S. Paris).

Titre — **Apport du système PASCAL à l'industrie.**

Ass. nation. Rech. tech., Inform. Document., Fr. (1972), no 2, 79-82. — Références de l'article

Résumé indicatif — Description générale du système PASCAL (Programme Appliqué à la Sélection et à la Compilation Automatique de la Littérature). Sources d'information : 15000 périodiques reçus à la bibliothèque du CNRS, analyses fournies par les centres de documentation sectoriels, brevets fournis par l'INPI. Traitements informatiques divers réalisés à partir d'une saisie unique : photocomposition du Bulletin Signalétique, tri des signalements, constitution d'index, constitution d'un stock documentaire, diffusion sélective. Produits : Bulletin signalétique, index annuels, bandes magnétiques, profils, recherches rétrospectives, reproductions. Dans le domaine industriel, une coopération avec divers organismes est réalisée pour l'électrotechnique et l'électronique, la métallurgie, la soudure, l'énergie, l'énergie nucléaire, les nuisances, les industries chimiques, les polymères, les transports.

C.N.R.S. Bulletin signalétique

Fig. 8

Conventional keyword indexes include synonyms - the aim of a semantically organized vocabulary is to eliminate synonyms. Accordingly, only about 2000 'pilot' keywords are included in the BRGM thesaurus.

The subject matter of the thesaurus is separated into semantic 'families'. Each family (class) is arranged to be as independent from the others as possible. The hierarchy within the classes is developed according to general usage and acceptance: out of the total of 45 families, 8 follow exactly the structure in other classification systems (sedimentary and crystalline rocks, stratigraphical age, palaeontology, etc.). The hierarchy of remaining families is arranged semantically in 'arrow diagrams' (flow charts). Such an arrow diagram is shown in Figure 9, which covers the same ground as the hierarchical classification example. The relatively small vocabulary and the graphic display of all keywords within a subject family facilitates the dissemination of information (indexing), while the arrow diagrams help in the organization of a search (retrieval) program.

Within each grid, the individual sub-subjects (keywords) are displayed in unique positions. The decimal grid number and the unique grid references (tens for rows, units for columns) are used for the machine coding of the keywords, and also for the numerical classification of the main themes of the publication. This dual use of the decimal grid allows the use of free-format input methods (see input techniques).

The hierarchical and semantic relations between keywords displayed in a grid are shown as unidirectional arrows, while the associative relations are indicated by bi-directional arrows. Capital letters indicate well defined and unambiguous terms, while small lettering indicates general terms. The relations between subjects in two separate grids are shown on the margin outside each grid.

For the upkeep of its complete geological and bibliographical documentation system, the BRGM uses a Bull Gamma 30 computer. Each month, eight booklets are produced, and these form eight simultaneous series, which list all the main Earth-sciences publications. An output example is shown in Figure 10.

There are plans to extend this system, and to make it the principal geological documentation system in France (i.e. merge with Bulletin Signaletique). Its possibilities are currently being evaluated by the International Council of Scientific Unions Abstracting Board, and by its Committee on Geological Documentation, created in April 1970. The aim of these organizations is to make geology the first scientific discipline with an internationally adopted multilingual thesaurus.

From this point of view, most of the member nations' systems are unsuitable for enborsement. The system used by the American Geological Institute bases the subject indexing on three-level term sets. The USGS, and also the Canadian Institute of Geology, work along the same lines. The disadvantages of this approach are: (1) the use of synonyms and, therefore, the subjectivity introduced into the indexing, and (2) retrospective search requires more than three levels. In the USSR, the leading documentation and abstracting agency is VINITI, which does not automate source data - a computer has been used only recently to prepare the author's index. Other large, or national systems, are either following these leads, or are using piecemeal solutions, where the principles of documentation and data storage are mixed up.

On the other hand, it is claimed that the decimal grid system has already demonstrated its ability to establish an identical indexing system in more than one natural language. A German-French thesaurus has been created: the German version of the BRGM system has been adopted by the Bundesanstalt fur Bodenforschung in Hanover. The Geofond geological documentation system in Prague (Czechoslovakia) is also modelled on this system.

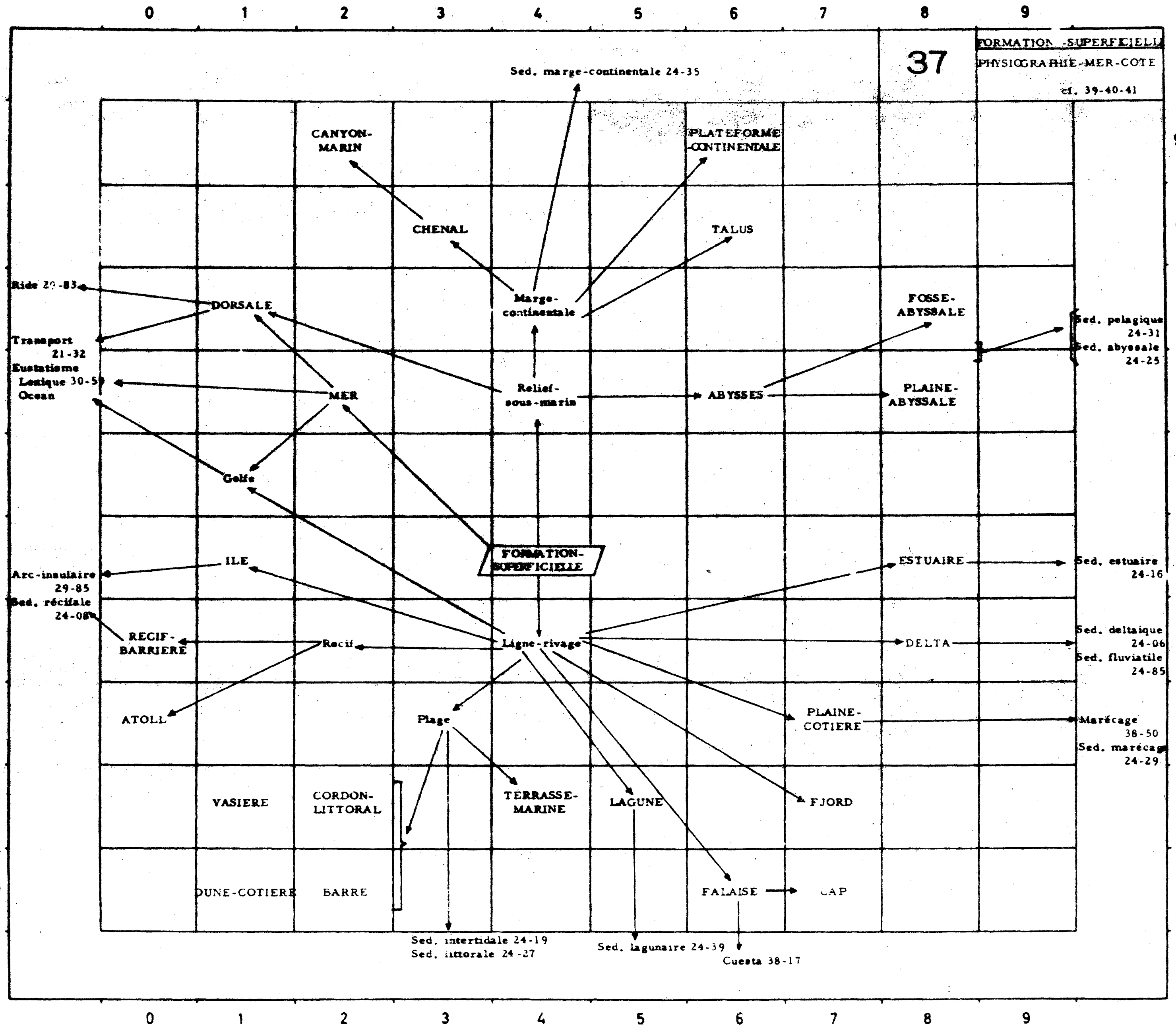
The easy indexing and grid retrieval methods, and its ability to overcome the language barrier are advantageous in this system. For the dissemination of information, the small vocabulary is also advantageous. However, the over-concentrated vocabulary is disadvantageous from the point of view of retrospective search: separate keywords for the various vertical, strike, cross, normal, reverse, etc. faults would be more desirable for this purpose than only the one keyword 'faults'.

(2) Retrospective search

The retrospective search ('flashback') facility of a system is the ability of that system to furnish - on command - references to a specific issue which might be only a minor part of a broader subject in several publications. Research workers currently working on or developing something new require such facility from an automated data processing system.

For the dissemination of information, the processing of the main subjects covered by a publication is sufficient; retrospective search requires

B.R.G.M.



FORMATION - SUPERFICIELLE

Fig. 9

30.000 références par an
items per year

accessibles par :
which can be retrieved by means of :

Section 224 - Cahier E

STRATIGRAPHIE ET GEOLOGIE REGIONALE

‡ 34 - 224 - 351. *BARR (F. T.), WEEGAR (A. A.)* **Stratigraphic nomenclature of the Sirte Basin, Libya.**
Tripoli, Pet. explor. soc. Libia, (1972), 1-179, 5 h.t., bibl. (2 p.), 51 ill. BRGM.
Au cours des travaux de prospection des petroliers de nombreux sondages ont ete faits et la connaissance de la stratigraphie de la Libye a considerablement progresse, mais a cause du caractere confidentiel de beaucoup de travaux, plusieurs systemes de nomenclatures paralleles s'etaient developpes. Ce livre est le resultat des travaux d'un comite charge de faire une homogeneisation. Des noms sont proposes pour quarante nouveaux groupes, formations et membres, reconnus en subsurface.

‡	34	224	351
l'article figure dans un autre cahier	annee de la collection	numero du cahier	numero de l'article
if cross referenced	volume	section	serial number

BARR (F.T.) — nom d'auteur / author's name

[- - - - - - - - - - - - -] — adresse de l'auteur ou du laboratoire / institute or private address

Stratigraphic nomenclature of the ...

TITRE DE L'ARTICLE EN FRANCAIS, ANGLAIS, ALLEMAND, ITALIEN, ESPAGNOL, NEERLANDAIS, PORTUGAIS, LANGUES SCANDINAVES avec traduction française pour toutes ces langues, à l'exception de l'anglais.
Traduction française avec indication de la langue d'origine pour les autres langues.
TITLE OF THE ARTICLE IN FRENCH, ENGLISH, GERMAN, SPANISH, ITALIAN, DUTCH, SCANDINAVIAN LANGUAGES
French translation for the other languages

Tripoli, Pet. explor. soc. Libia,	**(1972)**	**1-179,**	**bibl. (2 p.),**	**BRGM.**
nom de la revue, ou éditeur présenté sous forme abrégée / *abbreviated title of the periodical or publisher*	année / year	pages / pages	bibliographie / number of citations	bibliothèque / library

Au cours des travaux de prospection ... — résumé / *abstract*

- classification systématique des références
  classification of the citations and abstracts

- index matières
  indexes

● Monographie, Crétacé sup., Tertiaire, Nomenclature, Coupe sondage, Libye (Sirte), 351

— index matières permuté
permuted index

•Monographie, Cretace sup, Tertiaire, Nomenclature **Coupe sondage** Libye(Sirte), 351
•Permien, Siltstone, Gres, Dolomie, Faune, Coupe geologique **Coupe sondage** Sao Paulo, 382
•Methodologie, Lithofacies, Biotope **Mollusca** Analyse squelette, Fossilisation, 349
• **Monographie** Cretace sup, Tertiaire, Nomenclature, Coupe sondage, Libye(Sirte), 351
• **Monographie** Trias moyen, Dolomie, Microfacies, Diagenese, Sedimentation, Faune gastropode, Algae calcaire, Pelecypoda, Suisse(Tessin), 1[illegible]
•Jurassique, Cretace inf, Ammonitida **Monographie** Pakistan Ouest, 449
•Trias, Cargneule, Ecaille **Nappe** Suisse, 411
• **Nappe** Suisse, 579
•Monographie, Cretace sup, Tertiaire **Nomenclature** Coupe sondage, Libye(Sirte), 351
•Etude critique **Nomenclature** Primaire, Faune, Birmanie, 4[illegible]

— index géographique
geographical index

Libye
Monographie, Cretace sup, Tertiaire, Nomenclature, Coupe sondage, 351
Luxembourg
Domerien, Siltstone, Argile, Calcaire, Milieu marge continen-

— index matières spécialisé
subject headings

(en stratigraphie, minéralogie, paléontologie.).
(for some sections : mineralogy, stratigraphy, paleontology).

Cretace sup.
Calcaire, Marne, Faune mollusque, Madhya Pradesh, 450
Clastique, Calcaire, Flysch, Ph. Grece, 439
Clastique, Californie, 518
Couche rouge, Pamir, 435
Faune oursin, Pamir, 436
Nomenclature, Coupe sondage, Libye, 351
Nouvelle Castille, 45[illegible]
Radiolaria, Faune, Rocheuses, 445
Tertiaire
Calcaire, Flysch, Geosynclinal, Grece, 512
Calcaire, Flysch, Ride, Sillon, Iran, 4[illegible]
Monographie sommaire, Roumanie, 521
Nomenclature, Coupe sondage, Libye, 351

- index auteurs
  authors index

BALZER (H. J.), 558
BARR (F. T.), 351
BARTH (W.), 557
BASSET (A. M.), 585
BASSOMPIERRE (P.), 584

WALTHER (H. W.), 578
WATERLOT (B.), 582
WATKINSON (D.), 357
WEEGAR (A. A.), 351
WEIDMANN (M.), 142
WEINELT (W.), 573

Fig. 10

the retrievability of all minute issues dealt with in a publication. For this, the simple classification of each publication according to the main subjects of the title or the abstract is insufficient; all issues and aspects of the various subjects would need to be processed. This means the indexation of many well chosen keywords, some of which will only be used occasionally or perhaps never. Furthermore, each introduced keyword should be in a 'data state' (i.e. identified and uniquely referenced on input, and automatically connected to all relevant data items within the system). In achieving this, the basic issue is indexation at the time of input; the various input methods are examined from this point of view in Appendix 2.

Need for a new approach

Data processing requires accurate measurements, modulated, formulated, and critically evaluated data, completely objective descriptions, clear-cut definitions, sharply defined divisions, and distinctly separate classification categories. Not all geological information fulfils these requirements: 'there are not, and cannot be at the present time, many critically evaluated geological data, if by 'critically evaluated' is meant measurements which are repeatedly checked by specialists working with an ever increasing precision in different parts of the world'. (A. Hubaux, 2nd CODATA Conference, September, 1970).

Most geological objects cannot be adequately defined:

(1) Rocks of different composition progressively grade into one another. Their names imply more a certain mode of formation than exact composition, because it is often quicker to describe them genetically than to give a completely objective definition in the field. Geology is concerned with environment - therefore the interest of geologists in critically evaluated data and absolute standards are only marginal. The already established standards are more chemical, physical, thermodynamic, etc., standards than geological.

(2) The geological classification divisions follow the same genetic concept pattern. The three main classes of sedimentary, crystalline and metamorphic rocks are separated on the basis of origin, and not on composition. The different categories also grade into one another. Furthermore, in

describing the geology of a region, geologists make constant use of what they think to be the most probable history of a region. Ore deposits, faults, types of intrusions, etc., are all named on the basis of genetic concepts.

Genetically described geological data are accurate at the time of observation. It is a summary of all information available at the time; therefore, it is both useful and significant. However, at a later stage it might cease to be; geological knowledge is determined progressively, better and better, as more and more detailed information comes to hand.

Data processing on the other hand would require each data item to be critically evaluated and objectively described in the field. Only in this way can it be ensured that each data item is correctly placed in the classification structure at the first attempt for input, codification, and the purpose of information dissemination. In addition, only in this way can be built a reliable classification structure which does not immediately become out of date as a result of new information or research developments.

In a few geological disciplines (mineralogy, geochemistry, palaeontology) the establishment of a stable classification structure is possible. It is generally built according to a preconceived idea, on the basis of observations and previous experience. This is possible in these branches of geology, because the number of objects to be processed are either finite or close to being finite for all practical purposes, and their species are discrete (i.e. no chains of intermediate types exist). It is, therefore, useful and meaningful to study the properties of the species independently of their location and possible genesis, and to critically evaluate and objectively describe them. The question is: how should they be described?

In 1830, Lyell (Principles of Geology) stated: 'geology is the science which investigates the successive changes that have taken place in the organic and inorganic kingdoms of nature; it inquires into the cause of these changes and the influence which they have exerted in modifying the surface and structure of our planet'. According to this definition, geology is not a technical discipline, but an aspect of natural history.

The context and problems of geology were defined similarly by Bowen:

'Geological science is not concerned solely with processes, with the changes to which each material has been subjected. It seeks not only to know what earth substances are, but how they have come to be what they are, what factors have led to the observed arrangements and associations' (American Journal of Science, 1952). Bowen, therefore, proceeds to specify the features of the science mathematically.

Von Bubnov defines geology as a combination of the two (historical and mathematical) approaches: 'Quite correctly, geology is said to have the same objective as history in wishing to reconstruct a unique sequence of events. That reconstruction however is based on the creation of certain space time laws not on the individual, a radical distinction between geology and history. A single fossil is of interest only as a representative of a number of identical creatures, a biological reaction to geological processes, not an end in itself. Whether granite was intruded in France or Zululand is also all the same, just as in physics, it is of no consequence whether the refraction of light is observed on Jupiter or Saturn. Geological method does not individualise, it generalises ...' (Grundproblemen de Geologie Brontraeger, 1931).

These three definitions lead to the following conclusions:

(1) Geologists are interested in unknown processes which have developed in time. Conclusions drawn from them should follow logically from observations.

(2) The process is not known in advance, and the position of a sample station is often dictated by local conditions. Observation is carried out on random samples; therefore, observations are not connected with the processes.

(3) After the processes have been reconstructed from the study of their separate parts, the results are not connected to the history of the process. Basing their arguments on these conclusions, the proponents of mathematical geology see geology as a mathematical discipline. 'Thus I came to see that the mathematical tool for which I was seeking was one suitable to the description of nature, and I grew ever more aware that it was within nature iteself that I must seek the language and the problems of my mathematical investigations' (Wiener, 1956 - I am a Mathematician.).

The followers of Wiener believe that the advent of computers should revolutionize geology the same way as other sciences. They advocate the use of computers for the objective synthesis of accumulated information, and they suggest the abandonment of the traditional data collecting/ qualitative appraisal methods in favour of more vigorous quantitative methods (Merriam, 1965 - Geology and the computer, Geology Survey of Kansas.). Probability, stochastic modelling, and statistical and mathematical tools are used to solve selected geological problems (e.g. projective well log interpretation by Pierson). For this approach, the following definition is offered: 'Mathematical geology is a scientific discipline which deals with the establishment of mathematical models of geologic processes; geological processes are classified according to the type of stochastic process that - with a fundamental objective of investigating geology with mathematics - exhibit probability distribution functions with the necessary values chosen appropriately. All other applications of mathematics to geology, although they may have practical importance, are special cases or particular instances of solutions to problems that use mathematics in geology'. (Vistelius, 1968 - 23rd Session of the International Geological Congress, Prague).

Objective observation criteria is necessary for mathematical processing; in most fields of geology, however, the basic data are not characterized as such. The lack of objective determinations and the use of genetic descriptions reflect on the itemization, nomenclature, and also on the type of classification system used: mathematically these are indefinite and unstable owing to constant alterations and improvements. The mathematical approach, and its consequence: the fixed-format or hierarchical data processing systems are unsuitable in these fields of geology, because their first prerequisite is a stable classification system which is used in the input process, in codification, in the dissemination of information, and also in retrievals.

A new approach is necessary. A 'living classification' system would be the most useful in geology, one which is not created beforehand and which does not follow a preconceived idea, but which is created automatically, directly on the basis of data being fed into it. Only a flexible system could follow the continuous development in a natural science such as geology. The ideal system should be characterized by feed-back: a new item should either alter the automatically created structure, or fit into it completely. Such a system would need to give an individual consideration to each input item (allowing the use of free-format input methods), as well as to each characteristic (retrospective search requirement). The three main approaches which hold out a hope of achieving these aims are:

- classification by facettes
- inverted character processing

- semantics.

The first two are outlined in the Appendix; the semantic system, which has been chosen in France as a base for a geological data processing system on a national scale, is described below.

4. SEMANTIC DATA PROCESSING

PRINCIPLES

A language is used by man to express awareness in external reality. Influences are felt or observed; when experience is added to sensations 'information' is created, which can be communicated. The tools of communication are words, which can describe complex objects or complete concepts to others with similar experiences. Semantics is concerned with the meaning of words: it analyses the inherent contents of expressions, defines conceptual inferences, and takes into account those 'shorthand' elements, which allow people to transmit meanings to each other.

Geologists also communicate through words. Words in the geological language refer to individual objects (names), or compositional, formational, or genetical concepts (terms). Names and terms can transmit meanings between two similarly trained geologists, but they cannot establish a link between man and machines.

In most information systems artificial codes are used to interface human and machine languages. A computer can treat coded information: however, the application of a code requires many redundant operations in the input preparation stage. Texts need to be summarized, each information item formalized, classified, codified, and finally, recodified for machine use (expressed in binary form). Man needs to carry out all these non-productive, translating type of operations; therefore, the application of an artificial code requires an enormous amount of human interfacing at the data preparation stage.

In the semantic system those 'natural' processes which are being carried out by geologists every day as a matter of course are employed. When geologists examine the characteristics of a rock in order to name it, they are semantically 'encoding' it, and when they attribute a concept or meaning to a term, they are 'decoding' it. According to the semantic viewpoint, an information is already classified and coded when it is communicated; therefore, it does not need to be reclassified and recoded according to an 'artificial' code. Geological names, terms, and expressions are used in the semantic system only to recognize objects or concepts on input and trigger-off their semantic processing; this can be achieved without any subjective alterations to their form or to their original meaning.

The processing of the meanings of terms is possible in a semantic system because it employs the geological language as a code. Geologists use geological terms in conjunction with universally accepted classification systems. In a classification system, subject knowledge is summarized, and characteristic elements are distributed for easy reference as limits, partitions, subdivisions, and levels within a structure. Each delimited area within a classification structure is differently named - a name carries 'by definition' discrete information, infers identifying characteristics, implies theoretical considerations, and uniquely refers to a certain assemblage of information elements which characterize and define a concept. Geological communication is based on the common understanding and acceptance of these principles.

In the semantic system the defining elements, distinguishing characteristics, and, therefore, the meanings of geological terms are processed. The interfacing of words, names, terms, and expressions with their definitions and characteristic elements is automated in a 'machine dictionary', where each term is entered and its elements disseminated (uniquely addressed within the machine). The presence or absence of a distinguishing character or element is expressed in binary forms (1 for yes or 0 for no) within the 'semantic code', used for each term in the dictionary. This form of expressing meanings (1) satisfies both human and machine requirements, (2) establishes man-computer and computer-man interfaces, which can be tested for validity and checked for contents or meanings by non-computing geologists or computing non-geologists alike at any stage during the processing stages, and (3) uniquely identifies each geological term as well as each of its elements, and thus allows their in-depth processing (retrospective search facility for research work).

<u>The semantic code</u>

The basic tool of the automatic interfacing between the geological language on the one hand, and machine processing on the other is the 'look - up algorithm'. This is essentially a large matrix where columns are elements of a subject classifying checklist, and the rows are words in the geological language. A separate algorithm is used for the various rock types (one for all sedimentary rocks, another for igneous rocks, etc.). Each term in the geological language needs to be processed only once, however.

The checklist is a questionnaire which is designed to establish the inherent content of each geological term. Quantitative or qualitative properties are described by dividing the scale of variation into sections. Therefore, semantic attributes of geological terms are defined as:

$$a\ (x)\ b$$

where (x) is the value of the variable, and (a) and (b) are threshold values in the scale.

Questions referring to the columns of the algorithm may take the following forms:

"Is it red?" ; "Was it deposited in the sea?" "Did it have a detrital origin?" ; "Is there evidence of metamorphism?" ; "Does it contain Ca?" ; "Is there more than 10 percent quartz in the rock?" ; "Is the grainsize between 1/16 and 1/125 mm (silt size)?" ; "Does the rock contain at least 5 percent of modal olivine?" ; or "Is the ratio $\frac{Na_2O + K_2O}{Al_2O_3}$ greater than 1?"
Yes or no type of answers to these questions are shown as binary symbols (1 or 0) in the matrix. The matrix is entered in machine memory, where the symbols are equated to bits and the geological terms (rows) to a string of bits which constitute the semantic code. In the semantic code therefore:

(1) for each symbol, there is a corresponding word, word group, or phrase

(2) each symbol represents an answer to a conceptual question which is concerned with a geological element or a condition, the result of an observation, measurement, interpretation or conjecture

(3) the code acts as an interface from man to machine and also from machine to man, being self-explanatory for each geologist (through the questions) and capable of being expressed directly in machine code (binary symbols)

(4) the code conveys in total all the information necessary to describe, identify, and rigorously classify a geological object

(5) one or a combination of all the elements of information contained in the code can be used for sorting, listing, and higher processing

(6) the same code is used for synonyms, and for equivalent geological terms expressed in different languages, providing a base for international data exchanges

(7) the same code is used to reconstruct the original or standard name at the time of output, after the processing of all the various element combinations.

The number of pertinent questions (descriptors) required to uniquely distinguish a rock in a semantic code is surprisingly small; for example, 64 questions provide a unique semantic code for all igneous rocks. The machine requirement for the complete semantic coding of an igneous rock is one computer word (8 octets = 64 bits in an IBM system). Using this code, the computer is able to reply not only to questions about the word 'granite', but also to enquiries about rocks that contain quartz and feldspar, have a granular texture, contain no feldspathoids or olivine, etc.

The inclusion of geological terms in a semantic dictionary is almost unlimited. If 64 questions (or bits) are allocated for igneous rocks, the number of possible combinations is 2^{64}; on the other hand, there are only about two thousand igneous rock types, so there is room for terms which designate rock families, such as: 'rocks of variable composition ranging from granite to diorite'. It is also possible to code a group of words or terms into one term, or to select the questions (bits) for relevance. For example, if it has already been indicated that a rock can only belong to 'granites' as a group, a question about the ratio leucite/leucite + nepheline is not applicable.

Information levels

In a geological description, information can be found on various levels. A field name for a rock is less precise than the name given to that rock after a quick analysis, but this name might be changed again to a more precise one after a chemical analysis. 'Acid igneous rock' is more general than 'granite', which is in turn more general than 'biotite granite'. Descriptors of formations in an old well are less precise than those in a recently drilled well, and so on. In most fields of geology, the level of available information is important, as there is a difference between an observation (group of observed objects) and an interpretation (group of objects and ideas). Observations and incompletely defined 'low-level' information cannot be treated in most data processing systems currently in use, as these can only accept fully interpreted and clarified data.

In semantics, there are three levels on which geological information can be treated, and computer-stored concepts structured. The lowest level concerns the characteristics of a given rock, the next highest concerns the characteristics of a rock type or family of rock types, and the upper level is concerned with geological phenomena, theories, and past theories (concepts such as structural and tectonic stages, deposition, environment, geosynclines, and petrography formations); geological information which could be used for research on similar structures observed at various locations on the Earth and at various intervals of time; etc. The first level might be considered analogous to the characteristic properties of the elements in chemistry (valence, etc.); the second level to the presence or absence of C, H, O, etc., in the chemical formula; and the highest level to chemical radicals and grouping (Laffitte, 1969). Given this structure, a choice can be made in the level of programming according to the level of the task.

The structure may be further extended by the computer, either downwards to zero-level (where a distinction can be made between observation and interpretation, thereby identifying 'basic' geological concepts) or upwards to successively higher levels by dynamic classification of objects already processed ('living classification systems').

The machine coding process

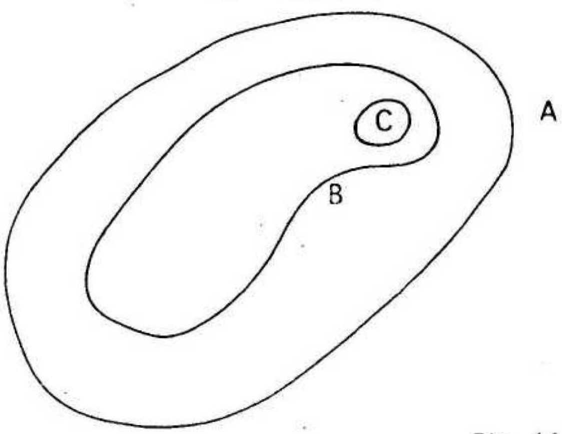
The semantic code is designed (1) to facilitate data preparation by allowing the recognition of geological terms and names in a written text; (2) to eliminate the constant need for a human interface in formalizing, classifying and coding input data item-by-item; (3) to disseminate information and to permit the processing of information elements; and (4) to store together original input items with their interpreted and implied meanings, and thus establish a facility for retrospective search.

The semantic coding process is based on a 'concept model' in which information levels, hierarchical relations, and general use are taken into account, and which is a collation of all compatible features of the various classification systems relevant to a subject. A separate model is needed for each subject in geology, and, in each of these, individual terms are defined conceptually and coded once, when the dictionary is established.

Where universally accepted hierarchical relations exist in a subject, these may also be expressed in the code. The dual hierarchical-conceptual character of the model and the code is illustrated in a mineralogical example below (reference: M. Kremer, École Nationale Supérieure des Mines de Paris), while a purely conceptual code for sedimentary rocks is outlined within the BRGM system.

Semantic classification - codification example: the processing of mineral names

Minerals as geological objects are well defined and classified according to their chemical composition and crystalline structure in the same way in several systems. They are subdivided into classes, subclasses, families, groups, species, and varieties. These hierarchical subdivisions are well known and universally accepted; the name "albite" automatically implies to a geologist a plagioclase group, a feldspar family, a tectosilicate subclass, and a silicate class. The first part of the semantic code for minerals, refer to those hierarchical relationships, where one mineral assembly as a whole is part of a greater group of minerals.



A = Assembly of sulphurs

B = Assembly of sulphosalts

C = Assembly of copper sulphides

(C< B)< A or : A implies B,
while B implies C.

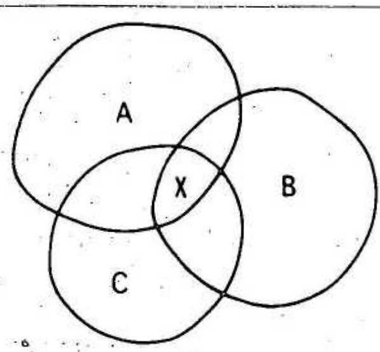
Fig. 11

Record 1975/9 M(G)570

These implied relations can be expressed more economically with a numerical code than with a conceptual code, so minerals in the first part of the semantic code are put into a list which is subdivided into four zones. The first zone represents class subdivisions (except for oxides and for sulphur), chemical families in the case of non-silicates, and structural families for silicates. The second zone subdivides the non-silicate isomorphic series, and the silicate groups and subgroups (amphiboles, garnets, etc). The third zone distinguishes between the mineral species, while the fourth zone between varieties; the sequential numbering of the species and varieties follow their respective family and species entries.

This arrangement allows coding short-cuts to be undertaken within the numerical part of the semantic code. In fixed-format systems each computer character expresses a single notion. In a conceptual semantic code a computer character can express eight notions using an algorithm (1 character = 1 octet = 8 bits), while here one zone can express up to 255 notions and is treated in the machine as a single character (1 zone = 1 octet). However, such coding short-cuts can be used only in subjects such as mineralogy where well known and accepted hierarchical relations exist.

The second part of the semantic code for minerals is used to express non-hierarchical relations, e.g. between all metals and semimetals, minerals of metamorphism or alteration, hydrothermal or opaque minerals, etc. In any other coding system, these mineral assemblages are delimited and classified separately in a way that none overlaps the others. In a conceptual coding system the various mineral assemblages can be made to intersect, and the intersections can be made to represent new assemblages:



A = Assembly of sulphurs

B = Assembly of all copper minerals

C = Assembly of all tin minerals

$X = A_{\wedge}B_{\wedge}C =$

- = sulphurs of tin and copper

Fig. 12

Record 1975/9 M(G) 571

These relations are expressed within the second part of the semantic code as a chain of bits, which also saves memory space (8 notions expressed by 8 bits within one computer character). Each bit represents and delimits a different mineral assembly; these are stored on discs, where each item is sequentially indexed and structured as follows:

(1) One blank octet (necessary for the sequential indexing).

(2) Mineral name: 29 octets (written in alphabetic mode, and acting as keywords).

(3) Semantic code : 10 octets

- numeric code : 4 octets (one octet per zone in binary representation)
- binary code : 6 octets (48 bits)

(4) Filter : 10 octets

The filter is equivalent to the pertinent bit notion (Fig. 13). If a bit in the semantic code (which might be 1 or 0) is pertinent, the corresponding bit in the filter is 1. On the other hand, if a bit in the code is not pertinent (if it does not make any difference whether it is 1 or 0), the corresponding bit in the filter is 0. By the use of the filter, a distinction can be made between minerals of a certain metal, and minerals which might contain that metal (for example, gold in pyrite).

Within the numerical zones of the semantic code, all bits are pertinent; therefore, the filter here consists entirely of '1'.

- numeric filter : 4 octets (when a numeric zone has a value, the corresponding 8 bits are all made to be 1

- binary filter : 6 octets

FAMILY	A	0010000000	1111				0010000000
GROUP	A B	0010100000	1111	1111			0010100000
SPECIES	A B C	0010100000	1111	1111	1111		0010100000
VARIETY	A B C D	0010101000	1111	1111	1111	1111	1111111111

CODE FILTER

Record 1975 9 Fig. 13 M(G) 567

Figure 14a is a simplified example illustrating the principles of semantic coding, whereas Figure 14b is an extract from the computer dictionary for minerals. In the printout, each keyword (mineral name) is shown against its code; the code octets are represented by their alphanumeric character equivalents. This printout format facilitates systems analysis work in the initial stages of building up a semantic system, as it provides a direct interface between the Friden typewriter (used in BRGM for the input of written texts through the automatically produced punched tapes) and the computer dictionary where the keywords are extracted from the text and their meanings are coded as binary symbols.

Grain-size attributes

1 Boulders
2 Cobbles
3 Pebbles
4 Coarse sand
5 Medium "
6 Fine "
7 Silt
8 Mud
9 Clay

Geological terms (Dictionary keywords, rows in algorithm matrix)	Part of the semantic code: grain-size attributes (columns of the algorithm)								
	1	2	3	4	5	6	7	8	9
ARENITE	0	0	0	1	1	1	0	0	0......
ARGILLITE	0	0	0	0	0	0	1	1	1
BOULDER BED	1	0	0	0	0	0	0	0	0
BOULDER CLAY	1	0	0	0	0	0	0	0	1
CLASTIC SEDIMENT	1	1	1	1	1	1	1	1	1
CLAY	0	0	0	0	0	0	0	0	1
CONGLOMERATE	0	1	1	0	0	0	0	0	0
COARSE CLASTIC SEDIMENT	1	1	1	1	1	0	0	0	0
COARSE SANDSTONE	0	0	0	1	0	0	0	0	0
FINE SANDSTONE	0	0	0	0	0	1	0	0	0
GREYWACKE	0	0	0	1	1	1	1	1	0
MUDSTONE	0	0	0	0	0	0	0	1	0
PEBBLY SANDSTONE	0	0	1	1	1	1	0	0	0
SANDSTONE	0	0	0	1	1	1	0	0	0
SHALE	0	0	0	0	0	0	0	0	1
SILTSTONE	0	0	0	0	0	0	1	0	0

Fig. 14/a

Simplified example of semantic coding

```
AIKINITE         =<12, ,3   >(2,18,30,12)*;
HUTCHINSONITE    =<12, ,4   >(2,30, 5,7)*;
PROUSTITE        =<12,8,1   >(2,7,5)*;
PYRARGYRITE      =<12,8,2   >(2,5,33)*;
MIARGYRITE       =<12,8,3   >(2,5,33)*;
ANDORITE         =<12,8,4   >(2,5,33,30)*;
FREIESLEBENITE   =<12,8,5   >(2,30,5,33)*;
PEARCEITE        =<12,9,1   >(2,7,5,16)*;
POLYBASITE       =<12,9,2   >(2,33,5,16)*;
STEPHANITE       =<12,9,3   >(2,33,5)*;
ARGYRODITE       =<12,9,4   >(2,5)*;
CANFIELDITE      =<12,9,5   >(2,5,37)*;
DUFRENOYSITE     =<12, ,5   >(2,7,30)*;
JORDANITE        =<12, ,6   >(2,7,30)*;
SEMSEYITE        =<12, ,7   >(2,30,33)*;
JAMESONITE       =<12, ,8   >(2,30,33,18)*;
BOULANGERITE     =<12, ,9   >(2,30,33)*;
COSALITE         =<12, ,10  >(2,30,12)*;
CYLINDRITE       =<12, ,11  >(2,30,33,37)*;
LORANDITE        =<12, ,12  >(2,7)*;
BERTHIERITE      =<12, ,13  >(2,33,18)*;
GLAUCODOT        =< 6,36,   >(2,7,18,14)*;
HALITE           =<15, ,1,  >(26)*;
SYLVITE          =<15, ,2   >(20)*;
CERARGYRITE      =<15, ,3   >(5)*;
BROMYRITE        =<16, ,1   >(5)*;
IODYRITE         =<17, ,1   >(5)*;
SALMIAC          =<15, ,4   >*;
CALOMEL          =<15, ,5   >(19)*;
FLUORINE         =<14, ,1   >(17,13)*;
SELLAITE         =<14, ,2   >(17,23)*;
ATACAMITE        =<15, ,6   >(4,16)*;
BOLEITE          =<15, ,7   >(30,16,4,5,3)*;
CARNALLITE       =<15, ,8   >(20,23,3)*;
CRYOLITE         =<14, ,3   >(17,26,6)*;
CUPRITE          =<18, ,1   >(1,16)*;
PERICLASE        =<18, ,2   >(1,23)*;
ZINCITE          =<18, ,3   >(1,44)*;
TENORITE         =<18, ,4   >(1,16)*;
MASSICOT         =<18, ,5   >(1,30)*;
SPINELLE         =<18,10,1  >(1,6,23)*;
HERCYNITE        =<18,10,2  >(1,18,6)*;
GAHNITE          =<18,10,3  >(1,6,44)*;
MAGNETITE        =<18,10,4  >(1,18)*;
FRANKLINITE      =<18,10,5  >(1,18,44,24)*;
CHROMITE         =<18,10,6  >(1,18,23,15)*;
MINIUM           =<18, ,6   >(1,30)*;
CHRYSOBERYL      =<18, ,7   >(1,6,11)*;
ARSENOLITE       =<18, ,8   >(1,7)*;
CLAUDETITE       =<18, ,9   >(1,7)*;
SENARMONTITE     =<18, ,10  >(1,33)*;
VALENTINITE      =<18, ,11  >(1,33)*;
PEROWSKITE       =<18, ,12  >(1,13,40)*;
CORINDON         =<18, ,13  >(1,6)*;
HEMATITE         =<18, ,14  >(1,18)*;
MAGHEMITE        =<18, ,15  >(1,18)*;
ILMENITE         =<18, ,16  >(1,18,40)*;
HAUSMANNITE      =<18, ,17  >(1,24)*;
SPINELLES        =<18,10>(1)*;
RUTILE           =<18, ,18  >(1,40)*;
ANATASE          =<18, ,19  >(1,40)*;
```

For the logical extension to this approach, see Appendix 2 (Keyboard-to-computer input methods). For recent adaptations, see Fletcher (1974).

The semantic information storage and retrieval system of BRGM

On 22.5.1944, the French Parliament passed an Act, which established a Common Mining Law in France ('Code Minier'). Articles 131-133 of the Code request the submission of detailed reports to the State from all subsurface operations which reach the depth of 10 m. BRGM was formed on 23.10.1959 to administer, document, analyse and distribute the enormous amount of geological, geophysical, and mining reports received as a result of the Act. Within ten years, information on more than a hundred thousand operations was summarized and released to the public to aid further exploration. In the initial stages, manual methods and conventional techniques were used to document information and to process data.

It was soon realized, that for research work and for the creation of national inventories the manual system is inadequate. In co-operation with the École Nationale Supérieure des Mines de Paris and under the direction of Professor Pierre Laffitte, work began in 1968 to lay down the fundamentals of a new type of data processing system where (1) all geological information is fed into independent subsystems which are compatible with each other and are all part of a greater system on a national scale; (2) duplications are eliminated and data is processed only once, preferably at the source; (3) all information preparing processes are automated; and (4) data treatment and retrieval is computerized.

The semantic data processing system, which developed as a result of this co-operation, allows the exchange of data between subsystems, as well as concurrent or combined processings. Well data, for example, can be treated on their own, within their own subsystem, or, in larger scale research projects, together with other data relevant to mining, geophysics, mapping, stratigraphy, structural geology, tectonics, palaeontology, petrology, geochemistry, geochronology, sample documentation (cores, cuttings, mineral species, etc.), results of analyses, log interpretations, petroleum technology, bibliographies, etc.; all within the same national system.

The semantic processing of geological information is based on standardized technical files in the BRGM. A different file format has been designed for the various types of subsurface operations (mining, geophysics, drillings excavations, speleology, etc.); each format is the basic document of a subsystem. The operating principles of all subsystems are the same -

only one is described in detail below, the well data/hydrogeological technical file.

A technical file serves a triple purpose: (1) as a source document for manual use; (2) as a direct input format (formalization and codification); and (3) as a base for a completely computerized data bank and data analysis system. When information is entered in the file, it is automatically prepared for computer input. This one-step input of data for both manual and computer use makes the BRGM system practical, time-saving and effective; it could provide a model for BMR, should a well data system be planned to form part of an Australian national geological data processing system. The BRGM system has numerous peculiarities, the result of different language, mapping, and administration systems in use in France. However, the system could be easily modified to suit Australia.

The well data/hydrogeological technical file

The well data/hydrogeological technical file consists of a double-sized printed format which when folded forms the front and back covers of the technical file (Figs. 15, 16). Several A4-size folios (Fig. 18). might be placed between the covers - each designed for the documentation of different types of specialist data (lithologic description, sample data, technical details, etc.). General bibliographical, technical, and stratigraphic data from oil wells, shafts, and water-bores are recorded in this file.

Most of the file is arranged as a questionnaire: the heading and titles are preprinted both on the covers and on the specialist folios. Details are filled in by a typewriter (Friden 1130) which prepares a punched tape at the same time. The completed cover and folio hard copies form the file which is available for manual use, while the punched tape is used directly for the input of all recorded data into a computer. This arrangement eliminates coding and card punching, and allows the visual verification of every input item.

Front cover General well information and bibliographic data are recorded on the front of the technical file. The following numbering refers to the numbered compartments shown in Figure 15.

0. Start of work (fixed-format field). In this blank space the following three symbols are typed: The correct typing and positioning of these symbols within the compartment is important, firstly, because it ensures that the paper is correctly positioned in the typing machine, and therefore, the tabulating arrows (see top of the form, Fig. 15.), the headings, and the corresponding data compartments are correctly placed, and secondly, because it marks the beginning of a new record on the punched tape and on the following magnetic

DÉPARTEMENT (1) .... N° D'ENREGISTREMENT (2) . / . .

COMMUNE (3)

CARTE GÉOLOGIQUE
N° (4) (5) 1/ . . .000
FEUILLE (6)

DÉSIGNATION (7) :

CARTE TOPOGRAPHIQUE (8) 1/ . . .000
FEUILLE (9)

OBJET (10)

Indice de classement (11)

DATE D'EXÉCUTION (12) . . / . . / . . . . à . . / . . / . . . .
APPROFONDI LE . . / (13) / . . . . à . . / . . / . . . .
PROFONDEUR FINALE (16) . . . . m
PROFONDEUR RÉELLE (17) . . m le . . / . . / . . . .

ARCHIVAGE : LETTRE DU . . / . . / . . . .
(14) MISE DANS LE DOMAINE PUBLIC LE (15) . . / . . / . . . .

NATURE (20)
MODE DE FORAGE (22)

COORDONNÉES LAMBERT
X : (18) . . . X' : . . . . . .
Y : . . . (19) Y' : . . . . . .
Z : (21) . . . Z' : . . . . . .
(23)
ZONE

MAITRE DE L'OEUVRE (24)
PROPRIÉTAIRE OU EXPLOITANT : (25)
N° CODE INSEE EXPLOITANT (26)
ENTREPRENEUR (27) :
TRAVAUX CONSEILLÉS OU SUIVI PAR (28) :
ORIGINE DES DOCUMENTS (29)

CROQUIS DÉTAILLÉ (37)

ÉCHANTILLONS (30) :

GÉOLOGIE (38)

CONTENU DU DOSSIER (31)

CONTENU DES ARCHIVES : (32)

ÉTAT DE L'OUVRAGE (33)
(34)

ACCÈS (35)

OBSERVATIONS (36) :

Mod. CMI/INF n° 1

REPÈRE ALTIMÉTRIQUE

DATE	NATURE DU REPÈRE	+ COTE DU REPÈRE * −	E.P.D. E.N.G. R.N.G.	CROQUIS DÉTAILLÉ DU REPÈRE
../../..		......•..		
../../..		......•..		
../../..		......•..		
../../..		......•..		
../../..		......•..		
../../..		......•..		

EAUX SOUTERRAINES

NAPPE CAPTEE

ÉTAGE GÉOLOGIQUE

N° code Agence de bassin :

NIVEAUX D'EAU OCCASIONNELS					PRODUCTIVITÉ DE L'OUVRAGE				
DATE	Profondeur du plan d'eau sous repère *	Hauteur d'ascendance au dessus du repère *	NATURE NIVEAU MESURE	+ COTE ABSOLUE DU PLAN D'EAU * −	DÉBIT EN m³/h. *	DURÉE DU POMPAGE * Heure Min.	Rabattement en m. en fin de pompage *	s t d	Température en °C
../../..	....•..	....•...	....	......•..	......•..	...H..	....•...		....•..
../../..	....•..	....•...	....	......•..	......•..	...H..	....•...		....•..
../../..	....•..	....•...	....	......•..	......•..	...H..	....•...		....•..
../../..	....•..	....•...	....	......•..	......•..	...H..	....•...		....•..
../../..	....•..	....•...	....	......•..	......•..	...H..	....•...		....•..
../../..	....•..	....•...	....	......•..	......•..	...H..	....•...		....•..

COUPE TECHNIQUE FORAGE

PROFONDEURS DE	A	⌀ intérieur en pouces	⌀ intérieur en mm.	OBSERVATIONS
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	

COUPE TECHNIQUE TUBAGE D'ÉQUIPEMENT

....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	
....•...	....•...	...•./.	....	

RÉFÉRENCES BIBLIOGRAPHIQUES

DOSSIER INSTRUIT PAR : .... CONTRÔLÉ PAR : ....

le ../../.... le ../../....

tape for computer use.

1. Department. Name of the administration subdivision of France (approximate equivalent of a shire). Names are treated as keywords; they are defined in the mapping subsystem.

2. Registration number. Reference to the registration system used is given in the first part as keywords, while the actual number is shown in the compartment. This way, there is provision to use different registration systems, without having to amalgamate them forcefully into a new universal system, and renumber them. (An adoption of this mode could mean automatic compatibility between State numbering systems and systems used by APEA, BMR, and private organizations).

3. Name of commune. Approximate equivalent of a County of Parish.

4. Geological map numbers.

5. Geological map scale.

6. Geological map name.

7. Designation. This is a free-format compartment, where any classification may be used. (In a BMR system this space could be used for the AAPG/APEA classification of oil wells, or for the nomenclature used in the Core and Cuttings Laboratory, or for any other terminology.)

8. Topographic map scale.

9. Topographic map name. Free-format, descriptive form.

10. Objective. Free-format, descriptive form.

11. File number. One numbering system is used in BRGM for all wells, bores, mines, shafts, excavations, caves, pits, and trenches. The same number applies to a well and to its technical file, thereby eliminating cross referencing and duplication.

This is a composite number, made up of three components: (1) 1:50 000 geological map number; (2) the number of the 1/8 subdivision of the above map; and (3) a unique well (file) number allocated to all works within the 1/8 area of the map. The three components are types in the three compartments shown in Figure 15.

BRGM regional offices (there are 11 in France, each covering a different 'Geological Province') control, make up, and allocate numbers for new works. A declaration concerning the new work is submitted to the local Chief Engineer, who forwards it to the BRGM regional office, where the exact well location is marked on the relevant 1:50 000 geological map. The marked position determines the 1/8 subdivision number on the map and identifies the number of the relevant topographic 'master map'. The consecutive number is determined and given to the new work.

The introduction of this type of well numbering in Australia would be a relatively simple matter, as such a system already exists in BMR, for all oil wells and BMR drillings. The unique well numbering system developed in the Core and Cuttings Laboratory is capable of considerable variation, and in each variation computers can be used to allocate, redistribute, or renumber wells in different map areas as proved by Esso (Fig. 17.). Locality codes (e.g. 10 second x 10 second, or any other unit area) could also be generated the same way using latitudes and longitudes.

12. Drilling dates. Spud and end of drilling dates are shown.

13. Extension dates. These are used when a drilling program is extended.

14. Confidentiality - archiving. A capital 'C' is typed if the file is on the confidential list, and 'D' if the well is released ('Domaine public'). The following date refers to the date of the relevant ministerial letter.

15. Release date. As determined by the Ministerial letter, referred to above.

16. Final depth is expressed in metres to two decimal places. If only an approximate depth is known, an asterisk (*) can be used, which will have two roles to play (1) it will carry the information that the depth shown is only approximate to the computer through the punched tape; and (2) in the manual use of the technical file, the repeat of the asterisk in the 'observations compartment' (numbered 36) will establish a cross-reference for a more elaborate remark (this is not stored in memory).

17. Real depth. The date of measurement is also shown here, e.g. the day an electrical log was run, which gave a different depth from the driller's depth. The real depth therefore indicates the effective depth of a certain date.

9:08   08-10-70   IJG   (9)

name of well	C & C Lab old number	new ESSO	TOPOG.SHEET	LAT.	LONG.
Dampier Downs No.1	55111/ 5	55155	3561	$18^{\circ}18'00''$	$123^{\circ}06'00''$
Doran No.1	55111/ 16	55155	3561	$18^{\circ}10'56''$	$123^{\circ}29'06''$
Edgar Range No.1	55111/ 15	55168	3660	$18^{\circ}45'26''$	$123^{\circ}35'33''$
Frome Rocks No.1	55111/ 2	55156	3661	$18^{\circ}11'48''$	$123^{\circ}38'42''$
Frome Rocks No.2	55111/ 3	55156	3661	$18^{\circ}15'15''$	$123^{\circ}39'35''$
Grant Range No.1	55111/ 6	55157	3761	$18^{\circ}01'00''$	$124^{\circ}00'25''$
Goldwyer No.1	55111/ 4	55155	3561	$18^{\circ}22'14''$	$123^{\circ}22'58''$
Matches Springs No.1	55111/ 17	55169	3760	$18^{\circ}41'27''$	$124^{\circ}03'12''$
Mt. Wynne No.1	55111/ 7	55157	3761	$18^{\circ}05'35''$	$124^{\circ}23'44''$
Mt. Wynne No.2	55111/ 8	55157	3761	$18^{\circ}05'35''$	$124^{\circ}23'44''$
Mt. Wynne No.3	55111/ 9	55157	3761	$18^{\circ}05'35''$	$124^{\circ}23'44''$
Mowla No.1	55111/ 17	55168	3660	$18^{\circ}43'50''$	$123^{\circ}42'35''$
Myroodah No.1	55111/ 11	55157	3761	$18^{\circ}16'15''$	$124^{\circ}11'27''$
Nerrima No.1	55111/ 10	55157	3761	$18^{\circ}26'55''$	$124^{\circ}22'17''$
Nerrima Bore	55111/ 12	55157	3761	$18^{\circ}28'16''$	$124^{\circ}24'02''$
Gogo Station Bore No.10	55112/ 11	55248	4061	$18^{\circ}17'00''$	$125^{\circ}36'00''$
Poole Range No.1	55112/ 3	55260	4060	$18^{\circ}53'06''$	$125^{\circ}41'20''$
Poole Range Oil test No.1	55112/ 4	55260	4060	$18^{\circ}53'06''$	$125^{\circ}47'20''$
Poole Range No.2	55112/ 5	55260	4060	$18^{\circ}52'27''$	$125^{\circ}49'02''$
Prices Creek No.1	55112/ 6	55260	4060	$18^{\circ}40'30''$	$125^{\circ}55'00''$
Prices Creek No.2	55112/ 7	55260	4060	$18^{\circ}40'40''$	$125^{\circ}55'55''$
Prices Creek No.4	55112/ 10	55260	4060	$18^{\circ}40'55''$	$125^{\circ}54'05''$
Prices Creek Bore	55112/ 2	55260	4060	$18^{\circ}42'00''$	$125^{\circ}53'00''$
Prices Creek No.3	55112/ 8	55260	4060	$18^{\circ}40'25''$	$125^{\circ}55'15''$
St. George Range Nol	55112/ 12	55171	3960	$18^{\circ}41'30''$	$125^{\circ}08'11''$

18-19. Location - X and Y co-ordinates (Lambert). Co-ordinates are expressed with 6 numbers, 3 to the left and 3 to the right of the decimal point. Unused positions are filled with zeros.

An asterisk shown in the compartments provided can mean that the co-ordinates are only approximate or that only the mean of several bores close to each other is shown. For manual use, a cross-reference to details in the compartment marked 'observations' (36) is provided.

Example: X : 735.228*

Y : 428.042*

Observations: *5 attempts were made to find water on the hospital grounds; the location of the only producer is shown.

If 5 wells were drilled and they all produce water for the same purpose (e.g. for the air-conditioning plant of a hospital) only the extreme co-ordinates are given;

Bore 1:	X = 361.08	Y = 80.00	Elevation: Z = +843.22
Bore 2:	X = 361.26	Y = 70.92	Z = +841.20
Bore 3:	X = 360.92	Y = 79.83	Z = + 841.31
Bore 4:	X = 360.72	Y = 80.10	Z = +842.70
Bore 5:	X = 361.15	Y = 80.00	Z = +842.15

. . . .

$X = 360.720 \qquad X^1 = 361.260$

$Y = 079.830 \qquad Y^1 = 080.100$

$Z = 841.200 \qquad Z^1 = 843.220$

20. Nature. The technical files of BRGM are so designed that the same format can be used to record geological data obtained from outcrops, road cuttings, canals, cliffs, trenches, quarries, mine pits, shafts, and open-cut workings, and from rotary, percussion, and auger drillings; to documents measured output of springs, artesian wells and water pump stations; to record natural cavities, caves, etc. In EDP, and also in the manual system, the various categories are identified in this compartment.

21. Elevation. Elevation or depth (Z) is expressed in metres, preceded by a '+' or '-' symbol for sea level reference. In the Z co-ordinate compartment, therefore, elevation is shown by a symbol, four characters left to the decimal point, two to the right, and a three-letter code:

EDP (indicating estimated height - read from a topographic map)

ENG (indicating estimated height - measured roughly)

RNG (for accurately measured heights).

These French mnemonic abbreviations could be replaced by KB, RT, etc., in Australia.

Examples: "Datum level elevation is 72.85 m above sea level, measured accurately.

Z: + 0072.85 RNG

For a series of wells, extremes are indicated as for the X and Y co-ordinates.

Example: for the example given in 19, and if the elevation is estimated:

Z: + 0841.20 EDP Z^1: + 0843.22 EDP

22. Drilling technique. Mode of drilling or excavation is indicated here, whether percussion, rotary or auger drilling; diamond coring; drilling with water or with air; subsurface sampling; dredging; blasting; etc.

23. Zone (Lambert). France covers 3 Lambert zones from north to south. These are coded: (1) for the north zone, (2) for the central zone, and (3) for the south zone. These zone boundaries are marked on all maps used in France.

Example: A well is drilled at X = 510.12 and Y = 198.17. Datum level elevation is 58.20 m above sea level, measured approximately from a bench mark. The area falls into the north zone.

X = 510.240 X^1 : ....

Y = 198.170 Y^1 : ....

Z = 0058.20 ENG Z^1 : ....

ZONE : 1

24. Title (or lease) holder.

25. Operator. Official company names or standard abbreviations are recognized as keywords, each tied to a full name and address in memory.

26. Company code (optional). The French Chamber of Commerce (INSEE) keeps a company register in which all registered firms are given a 12-character index number.

27. Financial partners.

28. Consultant(s). If consultants are not employed, the BRGM regional offices may provide free consulting services.

29. Origin of documents. BRGM regional services collected documents on old works from various public and private archives.

30. Sample details. Presence, absence, and whereabouts of samples are indicated here. The complete sample accession record is shown on a folio within the technical file.

31. File contents. The titles of all specialized folios, included within the covers of the technical file are listed here. Each entry points to a folio similar in type to either the front or back cover of the technical file (see also geological log folio example).

BRGM ensures that the folios are available first at the place of origin. There is a branch, section, group, or individual responsible for preparing a folio, placing the original into the technical file, and listing its addition on the cover. The punched tape is then forwarded for computer processing. Copies of the written folio are distributed as required.

File inserts special folios

The following is a non-exhaustive list of the possible specialist folios included in the well data/hydrogeological technical file:

a. Administrative folios

- Declaration card (summary)
- Declaration card (detailed)
- Accounting records of payments made or received. (In France, industry pays for the use of water in some areas. Receiving payments and carrying out studies to provide a sufficient amount of unpolluted water is also the responsibility of BRGM)

b. Geological data folios

- Regional geology
- Geological log (see folio example)
- Lithological descriptions
- Palaeontology
- Micropalaeontology.
- Petrology (results of microscopic examinations, x-ray, grainsize analysis, etc.)
- Isotopic age determination
- Chemical analysis of rock samples
- Geochemical analysis of rock samples
- Dips, stratigraphy, tectonics, etc.

c. Geophysical data folios

- Surface geophysical methods.
- Electric logs
  (Using a Benson pencil follower, logs are plotted on a uniform scale and put on magnetic tape, all in one operation. This arrangement is quick and cheap (about $5000). Running costs are about ½ cent per log/foot. Clerical or technical assistants can be used to trace the logs. This converts analog logs into digitized values. On output, logs are printed in a uniform scale, and placed in the file as folios (for manual correlations and for archiving), while the simultaneously produced magnetic tape is forwarded for computer processing (automatic correlations). Computer programs for formation recognitions and for correlations are available from Franlab (commercial annexe of the Institut Francais du Petrole).

d. Geotechnical data folios.

In each geological province a BRGM Regional Service Annexe consults and co-operates with industry on a commercial basis. Besides offering advice in the planning stages of new projects, the annexes also carry out studies in geological engineering, and publish geotechnical maps (Fig. 19). On these highly detailed maps, surface geology to the depth of about 20 m is shown (sources of economic substances such as lime, sand, gravel, etc.), together with

Example:

DEPTH (from)	DEPTH (to)	DIAMETER (ft,in)	DIAMETER ($^m/m$)	DESCRIPTION
-000.05	0000.00	18.0/0	....	Casing (protruding from ground)
0000.00	0024.00	18.0/0	....	Casing (cemented)
-000.30	0000.00	13.3/8	....	Casing (protruding from ground)
0000.00	0000.00	13.3/8	....	Casing (cemented, welded)
0000.00	0024.00	. /	....	Drill collar, cemented
0101.08	0269.00	09.5/8	....	Casing (cemented)
0250.08	0334.44	06.0/0	....	Perforated tube, filter

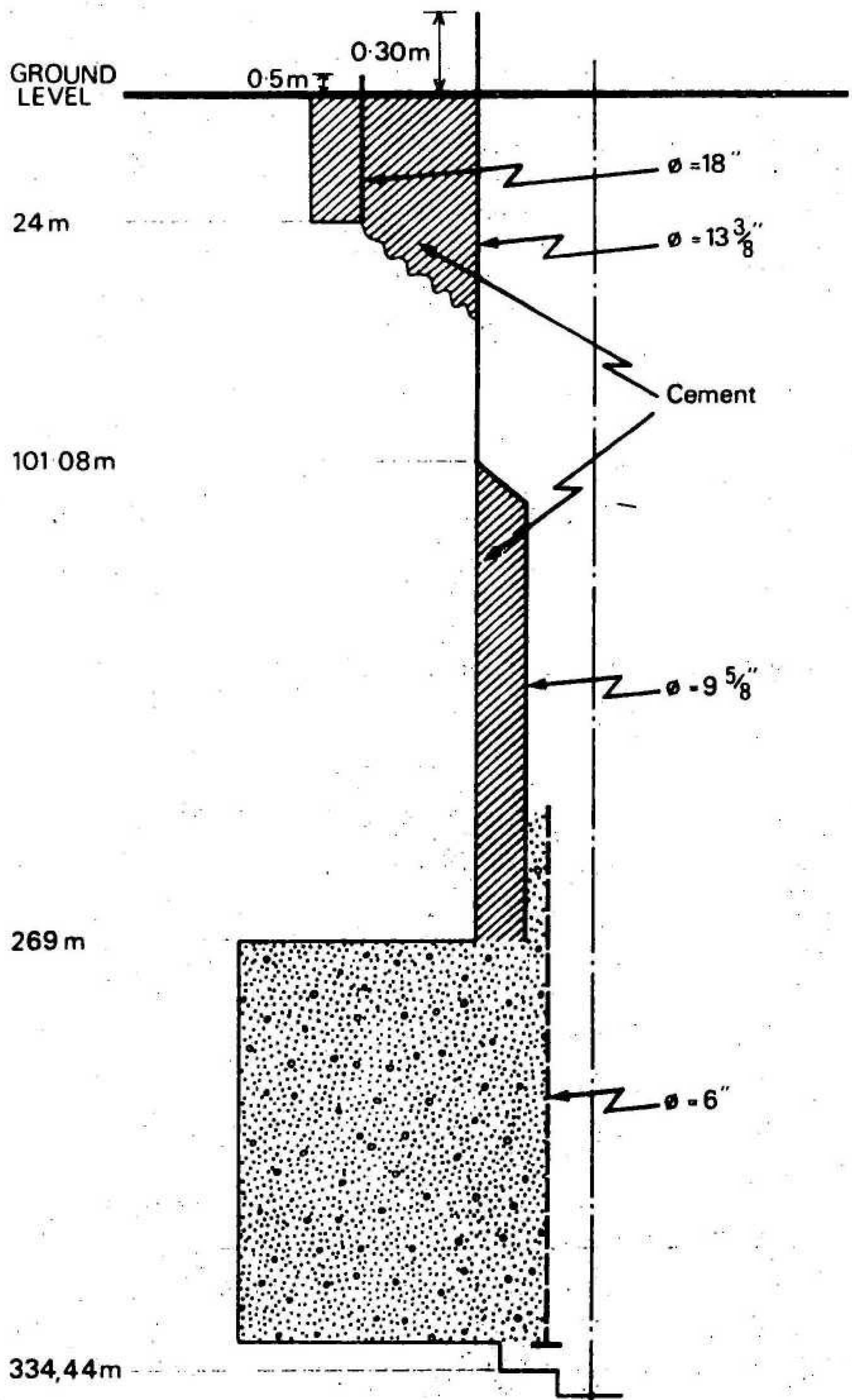
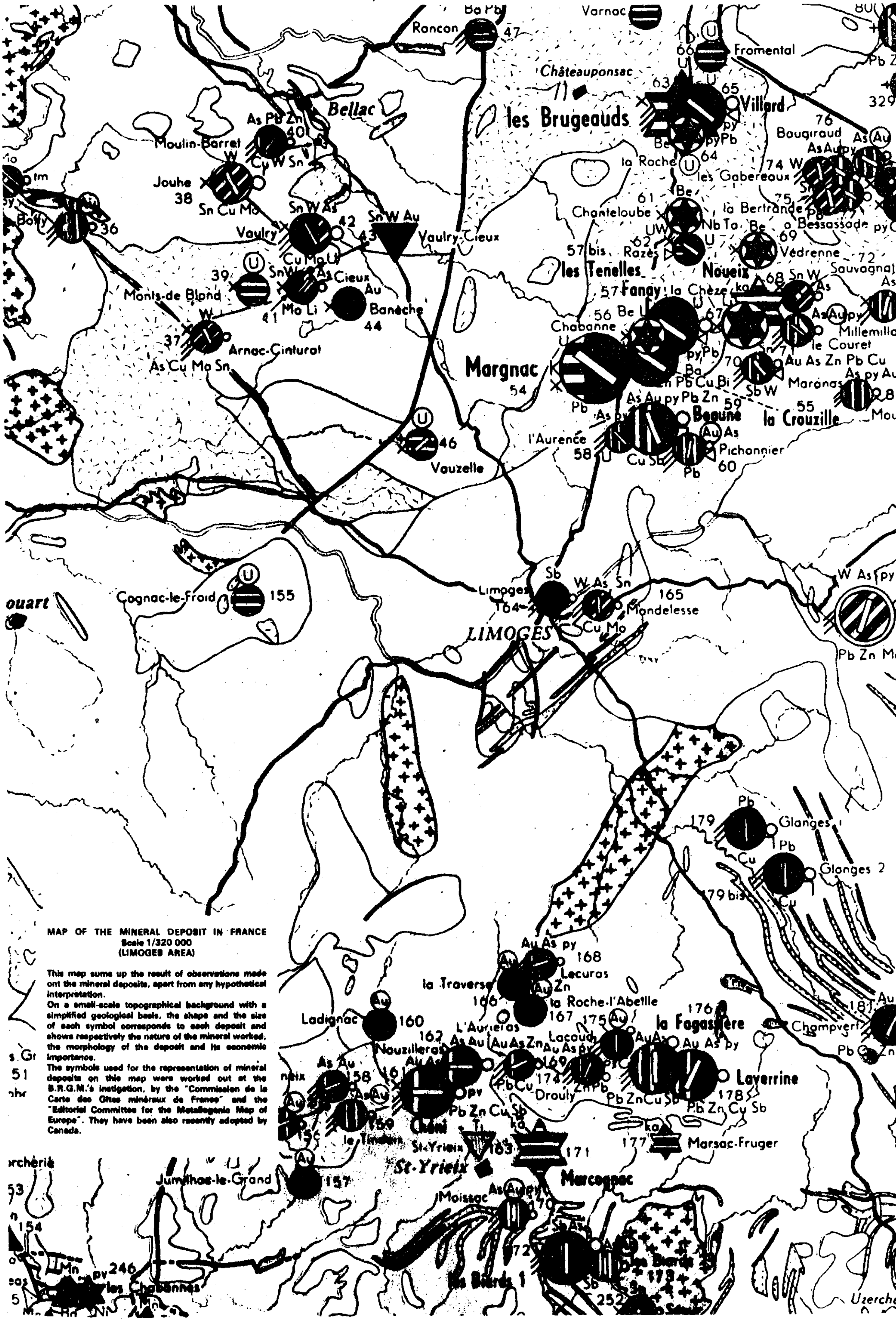


Fig. 18



MAP OF THE MINERAL DEPOSIT IN FRANCE
Scale 1/320 000
(LIMOGES AREA)
This map sums up the result of observations made ont the mineral deposits, apart from any hypothetical interpretation.
On a small-scale topographical background with a simplified geological basis, the shape and the size of each symbol corresponds to each deposit and shows respectively the nature of the mineral worked, the morphology of the deposit and its economic importance.
The symbols used for the representation of mineral deposits on this map were worked out at the B.R.G.M.'s instigation, by the "Commission de la Carte des Gîtes minéraux de France" and the "Editorial Committee for the Metallogenic Map of Europe". They have been also recently adopted by Canada.
Rancon
Varnac
Fromental
Châteauponsac
Bellac
les Brugeauds
Villard
Moulin-Barret
Jouhe
Bougiraud
la Roche
les Gabereaux
Chanteloube
la Bertrande
Bessassade
Vaulry
Vaulry-Cieux
Cieux
Rozès
Védrenne
Sauvagnat
Monts-de Blond
Banèche
les Tenelles
Noueix
Fanay
la Chèze
Chabanne
Millemilla
le Couret
Arnac-Cintural
Margnac
Maronas
Beaune
la Crouzille
l'Aurence
Pichonnier
Vauzelle
Cognac-le-Froid
Limoges
LIMOGES
Mandelesse
Glanges
Glanges 2
la Traverse
Lecuras
la Roche-l'Abeille
Ladignac
L'Auriéras
Lacaud
la Fagassière
Champvert
Nouzilleras
Laverrine
Drouly
Chéni
le Tindeix
St-Yrieix
Marsac-Fruger
Marcognac
Jumilhac-le-Grand
Moissac
les Biards 1
les Chabannes
Uzerche

Fig 19

hydrogeological data (such as limits and depths of known reservoirs, springs, rivers, catchment area, zones of protection, zones of overdevelopment, polluted areas, availability of resources, etc.). Data relevant to the production of geotechnical maps, and useful in the geological engineering consulting function of BRGM are recorded in this folio.

e. Hydrogeological data folios.

Basic hydrogeological data are recorded on the back cover of the well data/hydrogeological technical file (Fig. 16). The supplementary folios include: bacteriological analysis results; hydrochemical details; isotope dating; isotope tracing for water flow; pollution studies (spring and river pollution, reservoir contamination); reservoir performance studies (the BRGM decides whether or not a region can support more industry); piezometric studies (statistical water level measurements); hydrometric gauging; meteorological information; and speleological studies and measurements.

f. Technical folios.

Various technical details are listed in these folios, covering drilling techniques, drill stem testing, equipment used, exploitation details, etc. (Fig. 18).

g. Miscellaneous folios.

Bibliographies, plans, correspondence, progress and final reports, and evaluations are listed here, and are available in the master file (at BRGM) or in the Archives, but are not normally kept in open files.

32. Archival contents.

33. Accessibility. A summary and an explanation is given here of the present state of the working site, release date, and whereabouts of available information.

34. End of file. The end of record is marked "   " in this compartment (important for the paper tape, and, therefore, for computer use).

35. Availability. All comments in free, descriptive language.

36. Observations. All comments in free, descriptive language.

37. Plan drawing. All comments in free, descriptive language.

38. Geological summary. All comments in free, descriptive language.

Geological log folio

Geological information is recorded within the technical file on the geological log folio (Figs 23-25). The folio has four main headings: depth, lithological description, tectonics, and stratigraphy.

All terms used on the form in describing a rock or a formation (including even the tectonic, stratigraphical age, and fossil zone concepts) are recognized by the machine without further codifications, transliterations, etc., according to the semantic input principles.

Example: (Fig. 20) Between 20 and 50 m, the well penetrated fractured limestones, laminated marls, and injected, probably Jurassic clayey salt layer with brecciated gypsum at the base, and a fault plane. Below the fault, between 50 and 75 metres, 25° dips were observed in glauconite sand.

Depth	Lithology	Tectonics	Stratigraphy
0020.00	Limestone, white, granular Marl, spotted	Fractured Laminated	Bajocian Murchison zoone (*Cancellophyllus* fossil zone)
	Salt, clayey, redish, greenish Gypsum	Injected Brecciated	Probably Jurassic Toarcian and/or Aalenian
	x	Fault	
0050.00	Sand, glauconitic	D.25	Domerian
0075.00	....		

Fig. 20.

Besides the terms and concepts shown in the above simple example, and besides the formalized lithological description of all rocks, the following observed relations can also be expressed (and can, therefore, be recognized by the machine): discordances, discontinuities, variations in deposition, presence of distinct layers, progressive alterations (e.g. towards base), insertions, lenses, blocks, concretions, unit sizes and scales (in feet and inches or in metric measurements), probabilities, author's subjective preferences or speculations, etc. In filling up the geological log folio, and in communicating the above concepts straight to the computer, very simple rules have to be observed, and can be learned within a day by geologists and technicians.

(1) Depth. The depth of formation or unit tops is listed; the machine automatically works out thicknesses. Depth is expressed in metres to two decimal points. Unused characters are filled with zeros.

(2) Lithological descriptions. Two types of information 'elements' can be distinguished in lithological descriptions: A structural element such as the relative positions of and relations between two or more units (alternation, a superposition, interfingering, etc.), and lithofacies elements of the rocks themselves.

In the descriptions the elements written in normal geological language are recognized as keywords and treated as separate data items:

Example: Between 20 and 80 m, the well penetrated:

alternating	beds	(of) grey, compact limestone
(first data item with one term)	(second data item with one term)	(third data item with three terms)

brecciated in places	(and) yellow marls
(fourth data item with three terms)	(fifth data item with two terms)

(with) several sandy layers	(which are) uncemented (&) fossiliferous
(sixth data item with three terms)	(seventh data item with two terms).

Fig. 21.

Normal rules in writing descriptions need to be observed, however; main rock types should be written first, followed by quantitative and qualitative adjectives, colour, grainsize, sorting, etc., (limestones, compact, grey...).

Hierarchical ranking of certain elements within a description can also be indicated. Significant elements having a syntactic rank should be tabulated (typed in special places Figs 22, 24) to indicate to the machine the existing horizontal and vertical relations in the description.

<u>Example:</u>

Depth column	Lithological description column	Tectonics
0000.00	A————————	
	B————————	
	C————————	
	D————————	
	E————————	
	F————————	
XXXX.XX	G————————	

.... Fig. 22.

The positioning of significan t elements is determined by the tabulating marks (arrows) shown at the top in Figure 24, and the arrangements are governed by simple logical rules. In Figure 22 the vertical positioning of A and G shows an association of these two principal items within the same description; the precedence of A over G a superposition; B and F another association of two secondary items both included in A; C an insertion within B; F an interruption of that insertion; D and E an association within C; the precedence of D over E a superposition or perhaps a substitution of E instead of D; while in the sense of the horizontal hierarchy within the sentence, the positioning of A and its commencement in a higher order tabulation than B shows the order of importance. Figure 23 shows part of a conventional composite log; Figure 24 shows the formalized version of the same lithological description as it appears in the geological log folio; and Figure 25 shows a computer printout.

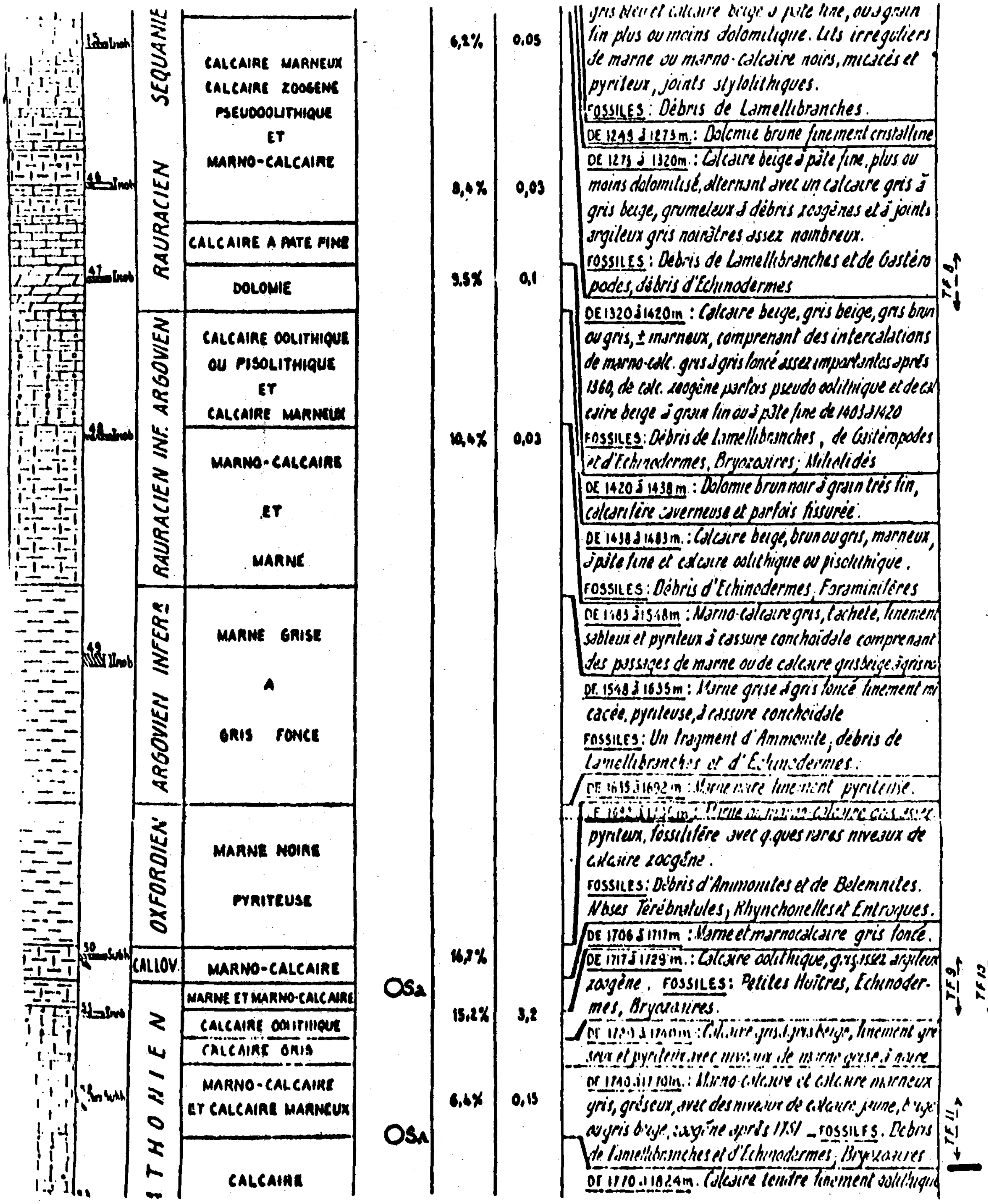

SEQUANIE
RAURACIEN
RAURACIEN INF. ARGOVIEN
ARGOVIEN INFER.
OXFORDIEN
CALLOV.
CALCAIRE MARNEUX
CALCAIRE ZOOGENE
PSEUDOOLITHIQUE
ET
MARNO-CALCAIRE
CALCAIRE A PATE FINE
DOLOMIE
CALCAIRE OOLITHIQUE
OU PISOLITHIQUE
ET
CALCAIRE MARNEUX
MARNO-CALCAIRE
ET
MARNE
MARNE GRISE
A
GRIS FONCE
MARNE NOIRE
PYRITEUSE
MARNO-CALCAIRE
MARNE ET MARNO-CALCAIRE
CALCAIRE OOLITHIQUE
CALCAIRE GRIS
MARNO-CALCAIRE
ET CALCAIRE MARNEUX
CALCAIRE
OSA
OSA
6,2% 0,05
8,4% 0,03
9,5% 0,1
10,4% 0,03
16,3%
15,2% 3,2
6,4% 0,15
fin plus ou moins dolomitique. Lits irreguliers de marne ou marno-calcaire noirs, micacés et pyriteux, joints stylolithiques.
FOSSILES : Débris de Lamellibranches.
DE 1249 à 1273 m. : Dolomie brune finement cristalline
DE 1273 à 1320m. : Calcaire beige à pâte fine, plus ou moins dolomitisé, alternant avec un calcaire gris à gris beige, grumeleux à débris zoogènes et à joints argileux gris noirâtres assez nombreux.
FOSSILES : Débris de Lamellibranches et de Gastéropodes, débris d'Echinodermes
DE 1320 à 1420m : Calcaire beige, gris beige, gris brun ou gris, ± marneux, comprenant des intercalations de marno-calc. gris à gris foncé assez importantes après 1360, de calc. zoogène parfois pseudo oolithique et de calcaire beige à grain fin ou à pâte fine de 1403 à 1420
FOSSILES : Débris de Lamellibranches, de Gastéropodes et d'Echinodermes, Bryozoaires; Miliolidés
DE 1420 à 1438 m : Dolomie brun noir à grain très fin, calcarifère caverneuse et parfois fissurée.
DE 1438 à 1483 m. : Calcaire beige, brun ou gris, marneux, à pâte fine et calcaire oolithique ou pisolithique.
FOSSILES : Débris d'Echinodermes, Foraminifères
DE 1483 à 1548m : Marno-calcaire gris, tacheté, finement sableux et pyriteux à cassure conchoïdale comprenant des passages de marne ou de calcaire gris beige à gris
DE 1548 à 1635m : Marne grise à gris foncé finement micacée, pyriteuse, à cassure conchoïdale
FOSSILES : Un fragment d'Ammonite, débris de Lamellibranches et d'Echinodermes.
pyriteux, fossilifère avec q.ques rares niveaux de calcaire zoogène.
FOSSILES : Débris d'Ammonites et de Belemnites. Nbses Térébratules, Rhynchonelles et Entroques.
DE 1706 à 1717m : Marne et marnocalcaire gris foncé.
DE 1717 à 1729 m. : Calcaire oolithique, gris assez argileux zoogène. FOSSILES : Petites Huîtres, Echinodermes, Bryozoaires.
gris, gréseux, avec des niveaux de calcaire jaune, beige ou gris beige, zoogène après 1751 FOSSILES. Débris de Lamellibranches et d'Echinodermes; Bryozoaires
Calcaire tendre finement oolithique
TF8
TF9
TF10
TF11

Fig. 23

Record 1975/9

de ment

PIÈCE : 1 FEUILLET : 3

signation :

upe établie par :

upe interprétée par :

Interprétation d'après : Échant. - Carte

X: . . . .

Y: . . . .

Z: . . . .

PROFONDEUR		NATURE DES TERRAINS Présentation / Description lithologique	TECTONIQUE	STRATIGRAPHIE
1248.00	—	Dolomie brune, cristalline		Kimméridgien
1273.00	+	Alternance de		
. . . .		Calcaire beige, dolomitisé, à pâte fine		
. . . .		Calcaire gris, beige, grumeleux, à débris zoogènes à		
. . . .		Joints argileux, gris, noirâtres		Kimméridgien, Séquanien
. . . .				
1320.00	+	Calcaire beige, gris, brun, marneux		
. . . .		Intercalation de		
. . . .		Calcaire argileux, gris		
. . . .		Calcaire zoogène, pseudoolithique		Séquanien, Rauracien
1403.00	—	Calcaire beige, à grain fin		///
1420.00	+	Dolomie brune, noire, calcarifère, caverneuse, à grain fin	fissurée	Rauracien inf.
1438.00	—	Ensemble de		
. . . .		Calcaire beige, brun, gris, marneux, fossilifère, à pâte fine		
. . . .		Calcaire oolithique, pisolithique, fossilifère		///
1483.00	+	Calcaire argileux, facheté, sableux, pyriteux, à cassure conchoïdale		
. . . .		Passées de		
. . . .		Marne		
. . . .		Calcaire gris, beige, noir		Argovien sup
1548.00	+	Marne grise, micacée, pyriteuse, fossilifère, à cassure conchoïdale		Argovien inf.
1635.00	+	Marne noire, pyriteuse		Oxfordien
1692.00	+	Ensemble de		
. . . .		Marne		
. . . .		Calcaire argileux, gris, pyriteux, fossilifère		
. . . .		Niveaux de calcaire zoogène	P.sub.	Callovien
1706.00	—	Ensemble de		
. . . .		Marne		
. . . .		Calcaire argileux, gris		Bathonien
1717.00	+	Calcaire oolithique, gris, argileux, zoogène		///
1729.00	—	Calcaire gris, beige, grèseux, pyriteux		
. . . .	—	Intercalation de niveau de marne grise, noire		///

Fig. 24

M (G) 596

```
0000 00        TERRE VEGETALE                                            QUAT.
0000 70        LIMON SABLEUX, BRUN, FIN                                  ///
0003 60        LIMON SABLEUX, JAUNE A GRAINS CRAYEUX                     ///
0004 50        LIMON ARGILEUX, SABLEUX, BARIOLE, GRIS, ROUX              ///
0005 50        ARGILE SABLEUSE, VERTE                                    LANDENIEN
0014 00        CRAIE BLANCHE, MORCELEE                                   SENONIEN
0017 00   FIN
*0038 3X 0176 01                                             01
      FORAGE EN BORDURE DU CD 117 RELIANT AYMERIES A PONT-SUR SAMBRE     707 365
      INTRAFOR-COFOR                                                     290 600
      ROSSIGNOL F.                                          XXX          +0129 00 EPD


0000 00        TERRE VEGETALE                                            QUAT.
0000 20        ARGILE JAUNATRE                                           ///
0003 80        SABLE GRIS, BLEU, ARGILEUX                                ///
0004 10        CALCAIRE GRIS, BLEU                          FISSURE      VISEEN
          9.90,18.10
               CALCITE                                                   ///
0020 00   FIN
      0038 4X 0194                                          01           01
      FORAGE AU LIEUDIT  FACHE DU CHEMIN DE DOMPIERRE                    712 195
      INTRAFOR-COFOR                                                     275 425
      ROSSIGNOL F.                                          XXX          +0167 00 EPD


0000 00        ARGILE SABLEUSE, GRISE, BRUNE                             QUAT.
0004 10        CALCAIRE GRIS, BLANCHATRE                    FISSURE      VISEEN
0005 85        ARGILE                                                    ///
0006 00        CALCAIRE GRIS, BLANCHATRE                    FISSURE      ///
0009 00        ARGILE                                                    ///
0009 30        CALCAIRE GRIS, BLANCHATRE                    FISSURE      ///
0013 00        ARGILE                                                    ///
0013 45        CALCAIRE GRIS, BLANCHATRE                    FISSURE      ///
0020 00   FIN
*0038 3X 0175 01                                            01
      FO*AGE EN BORDURE DE LA SAMBRE                                     707 975
      INTRAFOR-COFOR                                                     280 475
      ROSSIGNOL F.                                          XXX          +013. 00 EPD
```

Fig. 25

(3) Tectonics. Tectonic information is documented in the third column of the geological log folio. In this column also, words used in normal geological language are recognized, such as: fractured, faulted, concordant, diapiric, folded, anticlinal, etc. These keywords are already in the machine dictionary in the semantic system, together with their interpreted meanings. Each entry in the tectonic column is therefore tied to its machine code, and also to those depth, lithological, and stratigraphic entries, which are written on the same line of the geological log folio.

Example:

Depth	Lithology	Tectonics	Stratigraphy
0020.00	Limestone	Fractured	Permian
0050.00	Clay, salty	Injected	Triassic

Bearings and dips can also be expressed in this column. Bearings (azimuth direction) are given in degrees, preceeded by the letter A (azimuth), while dips are preceded by the letter P (pendage = dip). A 10° dip in the direction 45° is written as:

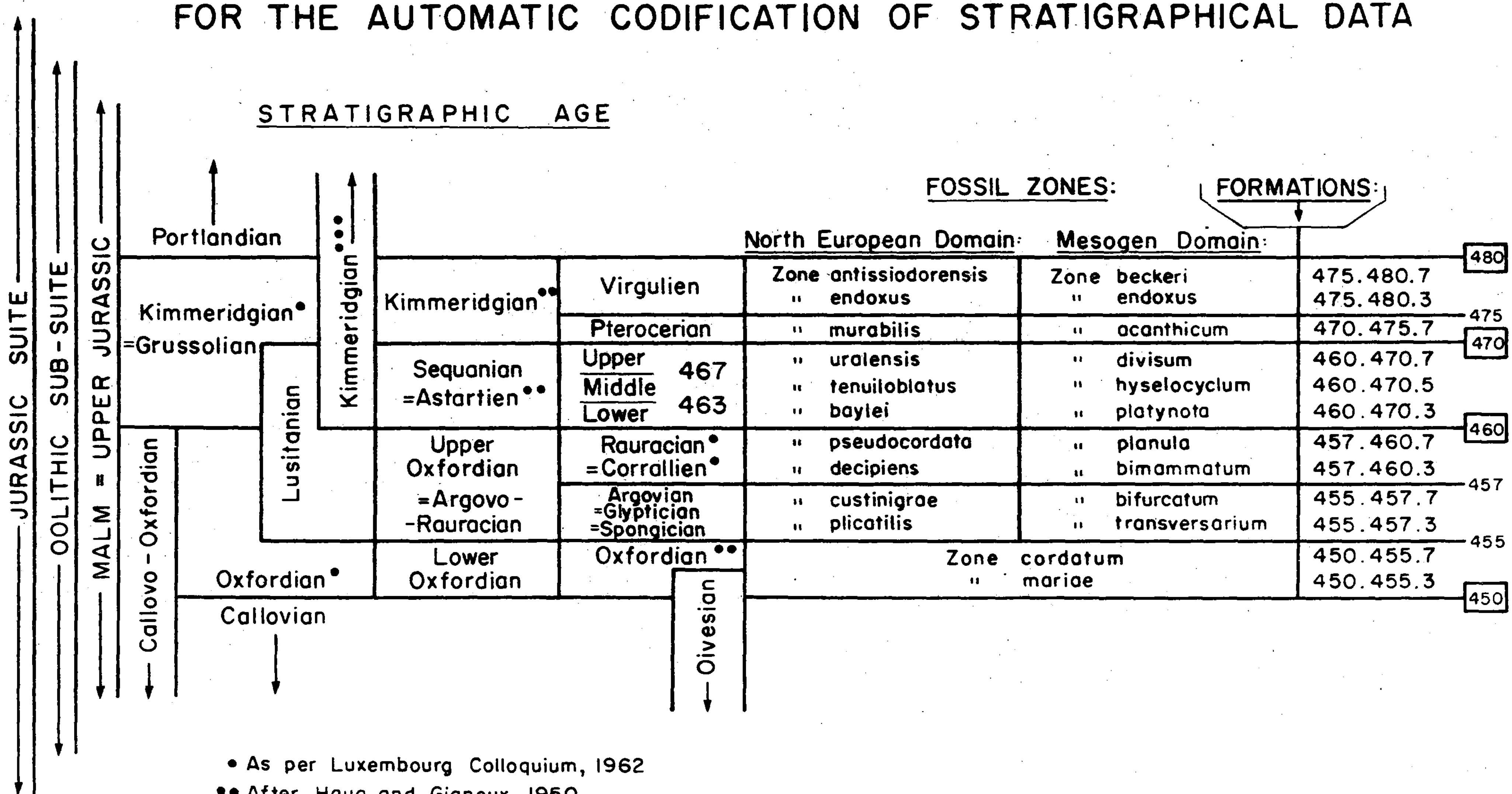
A.045 P.10

(4) Stratigraphy. The favourable setting of the Paris Basin allowed Cuvier and Brongniart to demonstrate at the turn of the nineteenth century that the basin was made of a great series of deposits laid down in a definite order, and that the various units may be traced by their rock-character and fossil contents over wide areas. The names of these units have now passed into international usage, and are being used as yardsticks for the determination of geological time in other places of the world where time lines and formation boundaries do not coincide. Geological time intervals established elsewhere are determined for each 'domain' (see Fig. 26) in France by periodic colloquia. The resulting uniformly accepted time and facies patterns allowed the development of a simple method of recording stratigraphical data in the semantic system.

Within the machine dictionary, the age interval limits of each input item (formation fossil zone, depositional duration, time interval) are determined and listed. For manual use, the list is presented as a master format, details of which are shown in Fig. 26.

When a stratigraphic keyword is fed into a computer, the dictionary equates it with its age code, shown on the right. The stratigraphic age code

DETAILS OF A SEMANTIC MACHINE DICTIONARY FOR THE AUTOMATIC CODIFICATION OF STRATIGRAPHICAL DATA



• As per Luxembourg Colloquium, 1962
•• After Haug and Gignoux, 1950
••• After Arkell

Fig. 26

Record 1975/9

B.M.R Petroleum Exploration Branch April 72 XM-3 is based on an arbitrary, numerical scale from 0 to 1000, which covers the whole geological history. The scale limits are consistent for each formation, fossil zone, depositional duration, etc., within each domain or locality. The determination of dichotomic relations between time lines and formation tops and their appropriate coding is the responsibility of the BRGM regional services. Formations already coded are listed on the master form the following way:

Standard name(s)	Local names.			
	Normandy:	Drome:	North (Calais):	570
Cenomanien				
= Gardonien				
= Pauletieu	Perche sand	Auriple sandstone	Tourtia formation	
	Maine sandstone			
				560

Fig. 27.

Where '=' signs are shown in the master form, the interval codes are the same for the equated terms; where the sign is not shown, the codes are different. At the appropriate domain or locality, the stratigraphic age code of a formation is allocated in the master form, against the scale.

<u>Example 1:</u> Montpellier Formation.

When the keyword 'Montpellier' is written in the stratigraphy column, the dictionary will allocate the numbers 715.735.6. The numbers 715.735 establish the top and bottom age limits, and automatically places the formation in the Sicilian depositional interval, within the Pleistocene Epoch (700-760) and within the Quaternary Period (700-800). The number 6 after the time interval indicates a marine environment.

<u>Example 2:</u> Zone Bidorsatum.

The use of the fossil name Bidorsata denotes a fossil zone (590.592) which determines a depositional duration of the Valdonian (also 590.592). This interval is automatically placed into the Campanian upper Senonian (590-595), within the Senonian Stage (580-598), within the Upper Cretaceous (560-600) of the Cretaceous period (500-600).

Example 3: Pre-Cenomanian tectonic phase.

The stratigraphic code for this tectonic phase is 560.560. The equal limits indicate the relatively short duration of the phase in the geological time scale.

Dichotomy, as well as doubt; can be shown in two ways:

(1) When there is uncertainty about the age of a formation, two or more units might be listed in the stratigraphy column. Retrieval will be affected under all of them. Similarly, when a formation transgresses a time line between two stratigraphic units, both units might be shown. If the position of the transgression zone becomes known at a later stage in the various domains, the correct unit will be allocated to the formation even if, originally, two units were shown in the stratigraphy column age filter in the dictionary).

(2) Different ages might be allocated to different layers, even within the same formation or sequence.

Example:

	Description	Tectonics	Stratigraphy
0020.00	Alternating		
	Limestone	Fissured	Lotharingian
	Marl	P40°	Lotharingian
	Clay, salty, red	Injected	Trias
0030.00	....		

5. A PROPOSAL FOR BMR

5.1 Presentation of the problem

Because there are at least a dozen service companies that could solve any single BMR data processing problem, and each would propose different techniques, input and output formats, and media to suit their own favoured hardware arrangement, a question must be asked: which would be the right one?

In deciding, there could be some pitfalls, which should be avoided. Some of these are listed below.

5.11 Single aspect systems. Short-term considerations in solving one problem at a time leads to an agglomeration of incompatible systems. 'Translating' (re-processing) an existing system to make it compatable with others in a larger system is possible, but this is an expensive undertaking.

A planning stage before any particular system is introduced should ensure compatibility with all the envisaged systems designed for other problems within the framework of larger systems (BMR, national, and international systems).

5.12 Input difficulties. Glossy leaflets and spectacular presentation of selected results can lead to overlooking input difficulties. When a system requires a complete translation and codification of all non-numerical and abstract informations, the following subsequent queries beg an answer:

- Who will undertake this job?
- If it is our present personnel, how can we re-educate them?
- Can we afford a considerable slow-down in output which would be the result of a general, enforced codification scheme?

A consideration given - not just on a relatively low and restricted level of what is happening to data once it is in a machine processable form, but also on the higher level of what needs to happen to data before this stage is reached - could eliminate this problem.

5.13 Personnel. Service, companies and advocates of 'package systems' are good at producing impressive lists of extractable results from their proposed system. The questions are here:

- How can these results be obtained?
- Do we have to educate each user to a programmer level, so that they know all the mechanics of data storage and management, input and output formats, equipment and software requirements to solve a problem, etc?

- Is the system user-oriented or computer-oriented?

5.2 General Solutions

5.21 Unquestionably, global considerations are required even when deciding on a solution to a relatively small and restricted data processing system.

Example: There are at least a hundred different ways in which the core and cuttings sample data storage and retrieval problem could be solved:

- manual, semi-automatic and automatic systems;
- punched cards, pre-punched cards, punched tape or magnetic tape;
- detachable store vignetting, magnetic character reading, mark-sensing, optical character reading, graphic character recognition;
- numerical systems, alphanumeric systems, codification, fixed-format, free-format systems;
- direct and/or indirect input and/or verification;
- instant display systems or error-checking by repunching, retyping or re-entry;
- visual displays for presentation or producing lists;
- manual or automatic distribution;
- or any combination of the possibilities listed above (the list is not exhaustive).

The cheapest proposal for data processing should not necessarily be chosen. A decision on a general principle concerning the type of network (i.e. hierachical, centralized, virtual channel, ring, etc.) is required first, followed by a set of system guidelines. The system should be discipline-oriented, covering all the data storage and retrieval requirements in BMR. There should be several data stores, all of which are interconnected on at least three levels (see 5.32), and the system should be capable of furnishing all information (name of formation, lithology, stratigraphy, analytical results - chemical, palaeontological, isotopic age, etc. - etc.) available on a sample.

5.22 The process of choosing a system should proceed, therefore, from the general to the particular, and not vice versa. For this reason, an outline of a possible ideal system is proposed below.

5.3 Outline of an ideal general system

5.31 Goals:

5.311: The design should aim for a general purpose and primarily problem-oriented system.

5.312: The system chosen to fulfill the design aim should be user-oriented and not computer-oriented.

5.313: The user language should be as natural as possible - programming experience should not be required to operate the system.

5.314: Users at all levels should be able to work with the named data, and the named data and program files, so that they are not required to know the mechanics of data storage management, input and output procedures, etc.

5.315: The data structure should have independent module sets in which each module has a direct access to all those resources and other modules it requires for solving a problem.

5.316: The system as such should be built on the principle of independent module sets in which each module has access to only those resources and other modules it requires for function. Apart from ensuring the confidential status of some data, this also means that package systems, which operate on the principle of enormous data banks where every data item is connected to every other, should be eliminated. Systems which are only devices designed to sell large computers with large memories should be recognized as such.

5.317: The system should be written in a high-level machine-independent language, as completely as possible; machine dependence in the implementation should be confined to as small portions as possible.

5.32 Operational levels. The system should be operational on three levels, using three distinctly different languages:

5.321: User-command Language - for use by non-computing, non-programmers. Access to the system should not be restricted to a selected few. This language should be completely machine and application independent.

5.322: Application language - to be used by programmers, for the purpose of inserting a new program into the system. All subsequent usage of the inserted programs should be governed by the user-command language, however. The application language should also be completely machine and application independent.

5.323: A system level language which embraces all details of the machine used and of the applications, and which allows access to programming and to data, storage, and management modules.

5.33 Data and information storage and retrieval. The user-command language as defined above is the one which most BMR personnel will use; the languages listed under 5.322 and 5.323 will be used mostly by computer professionals. For this reason, possible design aims and further details for this language are outlined below.

5.331: The user-command language should invoke processes which perform manipulations on a structure called a 'data state', which consists of a data item and all information associated with it.

5.332: With a data statement, which should control both the data to be manipulated and, implicitly, the sequencing of the process, the user-command language should be able to specify the data to be operated on.

5.333: The user-command language should be able to ask for any data item to be saved as a named file, or printed out.

5.334: There should be the facility to write macro-instructions in user command language, to combine any of the above functions.

5.34 Summary, user-command language. In general the user-command language, and the next highest - the application language (see 5.322) - should be interconnected in such a way that a program written in the application language can be named in the user-command language, and can be called at any time in the processing of data with interim or final parameters as desired. The complete data state (5.331) should be controlled by this principle, using user-command language. It should not be necessary to repeat in a routine operation an input-processing-output program written in Fortran or in another programming language each time, when a manipulation is necessary on data. In a data state, all these operations should be done automatically, so that the retrieval of results is governed by a simple user's language.

5.4 Conclusions

5.41: A system structure whose design is based on the capabilities of a particular computer; which is built on the concept of separate data files simply to facilitate the operation of the same computer; and which requires a program or subprogram each time a cross-referencing (a simultaneous consideration of data from two or more data files) is needed - can be operated only when a large number of programming personnel is available for carrying out routine operations and writing subprograms to satisfy individual demands. This modus operandi is cumbersome, the system is not responsive to providing up-to-date information, and there is the ever-present danger of an intended tool becoming the master.

5.42: The ideal system for BMR would be a user-oriented machine-independent geological and geophysical total system, in which during the processing of data, each data item is automatically cross-referenced to all other relevant data items (eg. established geophysical horizons to geological formations or beds, which are in turn tied to palaeontological and stratigraphical data, etc.).

5.43: There should be a three-tier language structure, to facilitate operations, to ensure confidentiality of some data, and to make the system approachable by programmers and non-programmers alike.

REFERENCES

BURK, C.F., Jr, 1973a - Computer-based storage and retrieval of geoscience information: bibliography 1970-72. Ottawa, Canad. Centre for Geosci. Data, Geol. Surv. Canad. Pap. 73-14.

BURK, C.F., Jr, 1973b - Computer-based storage and retrieval of geoscience information: supplement to bibliography 1970-72. Ottawa, Canad. Centre for Geosci, Data, Geol. Surv. Canad. Pap. 73-14 Suppl.

FLETCHER, J.G., 1974 - Characters in a dialogue. Datamation, 20(10), 42-47.

*HRUSKA, J., & BURK, C.F., Jr, 1971 - Computer-based storage and retrieval of geoscience information: bibliography 1946-69. Ottawa, Canad. Centre for Geosci. Data, Geol. Surv. Canad. Pap. 71-40.

LAFFITTE, P., 1969 - La codification sémantique en informatique géologique. Annales des Mines, 12, 75-83.

LAFFITTE, P., 1972 - TRAITEE D'INFORMATIQUE GÉOLOGIQUE. Paris, Masson & Sie.

ROBINSON, S.C., 1966 - Interim report of the committee on storage and retrieval of geological data in Canada. Nat. Adv. Comm. Res. geol. Sci., Geol. Surv. Canad. Pap. 66-43.

SOKAL, R.R., & SNEATH, P.H.A., 1963 - PRINCIPLES OF NUMERICAL TAXONOMY. San Francisco, W.H. Freeman & Co., 359 pp.

VICKERY, B.C., 1965 - TECHNIQUES MODERNES DE DOCUMENTATION. Paris, Dunod, 18 pp.

APPENDIX 1: CLASSIFICATION SYSTEMS

Classifications can be based on essentialism, nominalism, and empiricism; on each of these bases, object characteristics can be grouped together and arranged in a hierarchical structure.

Essentialism. A number of fundamental characteristics are selected and the objects are arranged in an hierarchical structure. Examples include the classification of living things and the classification of rocks. In the hierarchical structure, a characteristic determines each level; for example, the percentage of silica at the first level, the percentage of dark minerals at the second level, etc.

Nominalism. The names of related individual objects are arranged in an hierarchical scheme, according to groups, branches, classes, orders, families, genera and species. This concept is not used much in geology, but recently it began to play an important part in numerical taxonomy (Sokal & Sneath, 1963), where the individual items are constantly regrouped into definite areas according to the value of coefficients calculated from a large number of measurements. Numerical taxonomic methods are now used to produce worldwide inventories of various minerals (Ecole Nationale Superieure des Mines de Paris) or to process geochemical data (Centre de Recherches Petrographiques et Geochimiques in Nancy).

Empiricism. Items are arranged in an hierarchical structure of branches, families, etc. Subdivision is based on the maximum number of characters. For example, soils (Vickery, 1965) are classified by their origin in one hierarchy, and again in separate hierarchies by nature, structure, constitution, etc. Each soil type occupies a different place in each of the hierarchies (classification by facettes). With the advent of computers, empiricism is experiencing a revival also; using statistical, multivariance analyses, the principle is used in hydrogeology, trace element determinations. and pollution studies. The inverted data processing techniques used in semantics are similar in principle to the classification by facettes.

Classification methods.

In a classification system, the relation between the members of a class need to be established. In geology, ascending classification structures are built: affinities and common characteristics are recognized first, then the rocks are regrouped into classes. By contrast, palaeontologists attempt to establish a genealogical tree by placing together forms connected by descent. Both these types of classification systems are established after long periods of observing extrinsic characteristics, comparisons, and subjective reasonings, which, even so, do not alway lead to successful classifications; for example, at the present time, over 30 grainsize classification schemes are officially in existence in various countries.

Systems based on formalization by classification require that a particular classification be adopted. In a national system the classification must be (1) acceptable to all concerned (universities, research establishments, private companies, national organizations, etc.), (2) understandable (must not require a complete re-education of personnel), (3) enforceable (in France, an Act of Parliament gives powers to the BRGM to establish a national system), and (4) compatible with other national and international systems.

The following criteria could be used to determine whether a classification scheme is suited to ADP techniques: Is it suitable for source data automation? Is there a place for each item considered? Is the nomenclature unambiguous? A classification system should be easy to learn, simple to formalize, and should facilitate retrieval. There is no universal system with all these qualities suited to all disciplines. The reasons are briefly outlined below.

Dichotomic classification schemes have the advantage that they are simple for computerization, as each level can be treated as a disjunctive proposition (i.e. subdivided by 1 and 0). The level in the hierarchical structure is determined by the position of each individual unit in the code. The disadvantages prohibit the use of these schemes in a national system, however. They do not allow for widest possible uses, do not have the ability to answer a broad range of research type of questions, are unsuitable to create national inventories, and the codification necessary for automatic processing can become cumbersome very quickly. Single-handle retrievals are effected: a selection on the basis of two or more characteristics cannot be carried out simultaneously. Geological eras and periods, for example, cannot be placed in a dichotomic structure because several new words would have to be created to group together Pliocene with Miocene (Neogene?), Oligocene with Eocene (Paleogene?), etc; this would require a re-education of geologists.

Positional codification in hierarchical structures requires as many symbols as there are levels in the classification structure. The advantage of this method lies in the possibility of interrogating directly and at one level, by considering only the first, second, third, or fourth character in the code (IGS system).

The disadvantages are numerous, however. The coding list aims to be exhaustive to ensure an exclusive code for each item. This means that constant updating will be necessary as the science develops and new knowledge comes to light. Each change in the code will require the reprocessing of some of the data already in the data bank.

The code itself is not self explanatory, and its use is not simple. As each item needs to be manually codified before input, the bulky code book needs to be on everyone's table. For example, a petrologist is asked to identify a mineral in a thin section, look up the code for that mineral in the code book, write the code on a coding form, and so on for the next mineral. The form is then passed on for keypunching, verifying, etc. This way, formalization is achieved at the source, but there is no automation. The clerical work in data preparation is carried out by those whose principal occupation should be the using of a data bank for interpretations.

In addition, the forcing of all objects into a single hierarchical structure does not correspond to geological reality. For example, in the classification of rocks, grainsize distribution should play an important role, and it should be repeatedly considered at each level. This is not possible in a single hierarchical structure, where only one character is attributable to each level. The approach is, therefore, unattractive from the point of view of retrievals and research work.

Extended positional codification. Several attempts were made in France to extend the possibilities of the positional codification method, and to create a way by which two or more information characteristics can be coded simultaneously. Overlay filters and sequential codes superimposed over hierarchical or matrix arrangement achieved this, and resulted in less complex codes (a sequential code superimposed over a matrix is simpler to process than the subscripted variables of the matrix). However, data input remained difficult, and the creation of overlays presented difficulties. Furthermore, the resulting data bank did not allow direct interrogation about either one or the other characteristics. The extended positional codification method is, therefore, also unsuitable for research work, and for processing data on a national scale.

The Universal Decimal Classification Scheme aims to embrace all the fields of scientific activity including geology, and it claims to be open-ended (all developments are catered for by simple additions). The major international disagreements concerning the Universal Decimal Classification Scheme point to the fact, however, that the presently used system structure is strained and that it is kept going only by constant modifications, the results of more and more arbitary and precarious decisions. After almost a hundred years in operation, it is now admitted that the original simplicity has now deteriorated into a real complexity, in which layman's notions on logic, decimalization, development of a hierarchy, etc. are no longer applicable.

The Geosystems Decimal Classification sequence is part of a computerized information retrieval system (Geo-Archives). It is being developed by Lea Associates Limited, London, and it is used in three publications: (1) Geoscience Documentation (bi-monthly journal for geoscience information with papers, reference sources, documentation and policy studies, review locator, and 'Geoserials' - the world list of geoscience serials), (2) Geocom Bulletin (the monthly information and abstract on mathematical and computer methods in geoscience, including 'Geocom Program' supplements and computer contributions), and (3) Geotitles Weekly (current awareness service).

The decimal classification scheme uses standard formats for the storage and presentation of bibliographical information. Within each class, the sequence is determined by the related fields of interest of the items; this helps to bring together for example bibliographies of soil mechanics (1028) and sedimentary processes (1033). A guide for the use of this system is shown in Figure 28. The standard formats used by Geo-Archives are in the process of being made compatible with other systems in Britain.

Classifications by facettes. Unlike UDC, the systems which are based on classifying objects by facettes intentionally abandon all pretentions concerning universality. In the system of facettes by Vickery (1965), and also in the one developed by the Classifications Research Group in the United Kingdom, the accent is on considering an object, a text, or a document simultaneously in various ways, which means that linearity, as well as systematic development of hierarchy, are also abandoned in these systems. The possibility of considering several aspects together allows for very fine distinctions to be made between two or more object combinations, and this advantage is used in these systems.

Example: In the main subdivisions in the English System, under the headings of 'Soil Science' or 'Pedology' for example, 350 terms are listed:

- 'mud: partially decomposed organic material'
- 'Cultivation: work affected on the soil to increase productivity'
- 'plough: equipment used for cultivation', etc.

Each of these terms is listed in 18 categories for soils, for example:

- 'soil, in terms of origin'
- 'soil, in terms of inhibiting organisms'
- 'operations carried out on the soil', etc.

Within each category, terms can be listed hierarchically, if this is found advantageous; for example, in the category of 'soils in terms of climate', these are the possibilities:

- arctic
- temperate
- subtropical and tropical: lateritic, red earth, etc.
- humid, etc.

In describing an object, several categories can be used, one term per category.

<u>Example:</u> The concept of 'bacterial weathering of laterites after ploughing' is described in the system by terms chosen from the following categories:

- 'soil in terms of climate'
- 'soil in terms of origin'
- 'soil in terms of inhibiting organisms'
- 'operations carried out on the soil'
- 'equipments used in operations'

etc.

Very detailed and very fine descriptions of concepts can be achieved by using this system. The original system was made to be exhaustive, however, and it is in this respect that the system fails. Only constant updates can keep such a system operational because of the continuous developments. Therefore, the French view is that, for a national system, all chanelling of informations, hierarchical structures, as well as all pretentions concerning exhaustivity, should be abandoned.

<u>Inverse data processing and automatic classification system.</u> Testing is a basic activity in the sciences: it is a confrontation of a preconceived idea, hypothesis, theory or law with data, and is the result of an observation or measurement. In the exact sciences (physics, mathematics) the results of tests conform to and confirm the laws of theories.

Geotitles Weekly arrangement

THERE ARE THREE PARTS TO EACH ISSUE OF GEOTITLES WEEKLY

1 YELLOW PAGES FOR NEWS

INFORMATION	ORGANISATIONS
PUBLICATIONS	EQUIPMENT
CONFERENCES	SERVICES

2 WHITE PAGES FOR SUBJECT CLASSIFICATION

ARRANGEMENT: CLASSIFICATION AUTHOR
Title
SOURCE

EXAMPLE: 421(430:116.2):332:552:740*30 DENGLER, H et al
The marine sedimentary iron ores in the Upper Jurassic [Malm] in northwest Germany BGJAH(79)69p214

CLASSIFICATION: By Geosystems Decimal Classification (GDC)

Subject field[1]	421	Iron & manganese ores
Geographical location[2]	(430)	Germany
Stratigraphic age[2]	(116.2)	Jurassic
	:332	Exploration geology
Related subject fields[1]	:552	Sedimentary rocks
	:740	Historical geology
Language[2]	*30	German

[1] Complete list on back cover
[2] By Universal Decimal Classification (UDC)—*see inside back cover*

AUTHOR: One or two authors are indexed; in the case of three or more authors, only the first is given, followed by et al

TITLE: Titles are given in English

CROSS REFERENCES: # CROSS REFERENCES are given at the end of sections. The main entry is found by turning to the section following the # sign:

332 # 421(430:116.2): 332:552:740*30 DENGLER, H et al
The main entry is found in section 421 Iron & manganese ores

SOURCE: Source journals are coded with a five letter GCODE:

GCODE(volume/number) year page
and indexed weekly on the green pages

3 GREEN PAGES FOR INDEXES

SOURCE INDEX:

BERIU (47/3)69	BULLETIN of the EARTHQUAKE RESEARCH INSTITUTE, UNIVERSITY of TOKYO
BGJAH (79)69	BEIHEFTE zum GEOLOGISCHEN JAHRBUCH [Hanover]
BMOIG(45/1)70	BYULLETIN MOSKOVSKOGO OBSCHESTVA ISPYTATELEI PRIRODY, OTDEL. GEOLOGICHESKII

AUTHOR INDEX:

DEL BONO, G & PIERUCCINI, U	936(45):710:360*50
DELLWIG, F	860:912
DENGLER, H et al	421(430:116.2):332:552:740*30
DENNE, WA	515:720
DENNINGER, E	750:595:815

The geological processes are not known in advance, however. The observations and field measurements are used to make up a hypothesis, while the testing is carried out to support or refute that hypothesis. Many potentially valuable results are being 'shelved' or 'doctored', simply because they do not support that particular hypothesis which is being applied. They would support the right one - but which is the right hypothesis? The question need not be asked if the hypothesis is built on the basis of the data and nothing but data, and if the models could be made to follow the available data instead of the other way around.

The advantages of the 'inverted' character processing method have already been proved in general geology, hydrology, trace element determinations, pollution studies, etc., in France. For the purposes of an illustration, a biological example is outlined below, because it is a general one, and because it is fully documented in the English language (Margaret O. Dayhoff, 1969 - Atlas of protein sequence and structure. Vol. 1,1966; Vol. 4).

For centuries, biologists tried to establish the geological 'tree of life', using morphological features, characteristic developments during the embryonic stage, and palaeontological evidences to distinguish between the individual species and to establish the relations between them. Neither of these approaches proved to be universally useful, as it could not give exact answers to questions concerning parentages, relations, evolutionary mechanisms, etc. These questions are now being answered biochemically, in the new science of palaeogenetics.

Palaeogenetics is concerned with proteins, which determine the structure and functioning of each living organism. They are important in life, as proteins form the different tissues, execute and control physiological operations, and ensure the defence of the organism. In each species a different protein assemblage is present which operates differently, and in each individual a different set of heredity features is acquired, which establishes identity. The amino-acid sequences in proteins, therefore, determine the place of an individual within a species and of each species within an arborescent evolutionary structure: the phylogenetical tree.

In the inverted data processing of protein sequence characteristics computers were used to establish a genealogical tree automatically, entirely on the basis of protein data. In each protein, the amino-acid sequence has a length of 110 links. Each pair of the sequence is compared position by position, and the number of amino-acid differences tallied. The sequence differences are arranged in a matrix which shows, for example, that the only difference between the man and a rhesus monkey is a group of amino-acids which is represented by the 66th position in the sequence. There are 9 positional differences between the man and the rabbit, and 12 between the man and the horse. But the rabbit and the horse differs in 6 positions, not 3: these are the differences which establish the various common ancestors.

If the three sequences are compared, 7 positions (19, 20, 23, 54, 58. 66, 91) are different between the man on the one hand, and the rabbit and the horse on the other; in four positions (55, 68, 70, 100) the man and the rabbit both differ from the horse; in one (52) position the man and the horse have the same amino-acid group, but the rabbit's is different; and in one position (97) all 3 have a different group. The statistical modelling of mutations are used to evaluate automatically these differences, and to establish the positions of the various common ancestors in the phylogenetic tree.

Three computer programs are used for this purpose. The first program is a fundamental one which is used in both of the others. Its aim is to establish topologies, and to connect the sequences according to available data. One-by-one all amino-acid positions are examined along the chains: when a single amino-acid recurs in all the sequences, it is treated as an original group. When a group is not present in all the sequences, a 'point of earliest time' is established on the phylogenetic tree. The program thus makes inferences concerning 'ancestral sequences' for each branch. By calculating the number of changes (mutations), the first program establishes the relative length of the branches and the height of the tree. The resulting relative chronology is later-on transferred into absolute terms by ties to geological times.

The second program follows the first, and it establishes the topology of the tree. It starts with three items, which are taken to represent a junction and three branches (Fig. 29a). The program then takes a fourth, and tries to fit it successively into three possible places (Fig. 29b. c. d). Each time, using the first program, it evaluates the topology of the four branches thus obtained, and chooses the best (the only possible) fit.

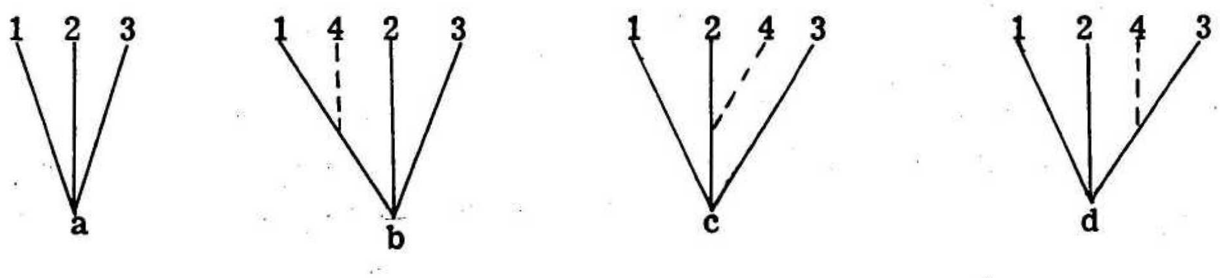


Fig. 29

THE PHYLOGENETIC TREE OF MAN

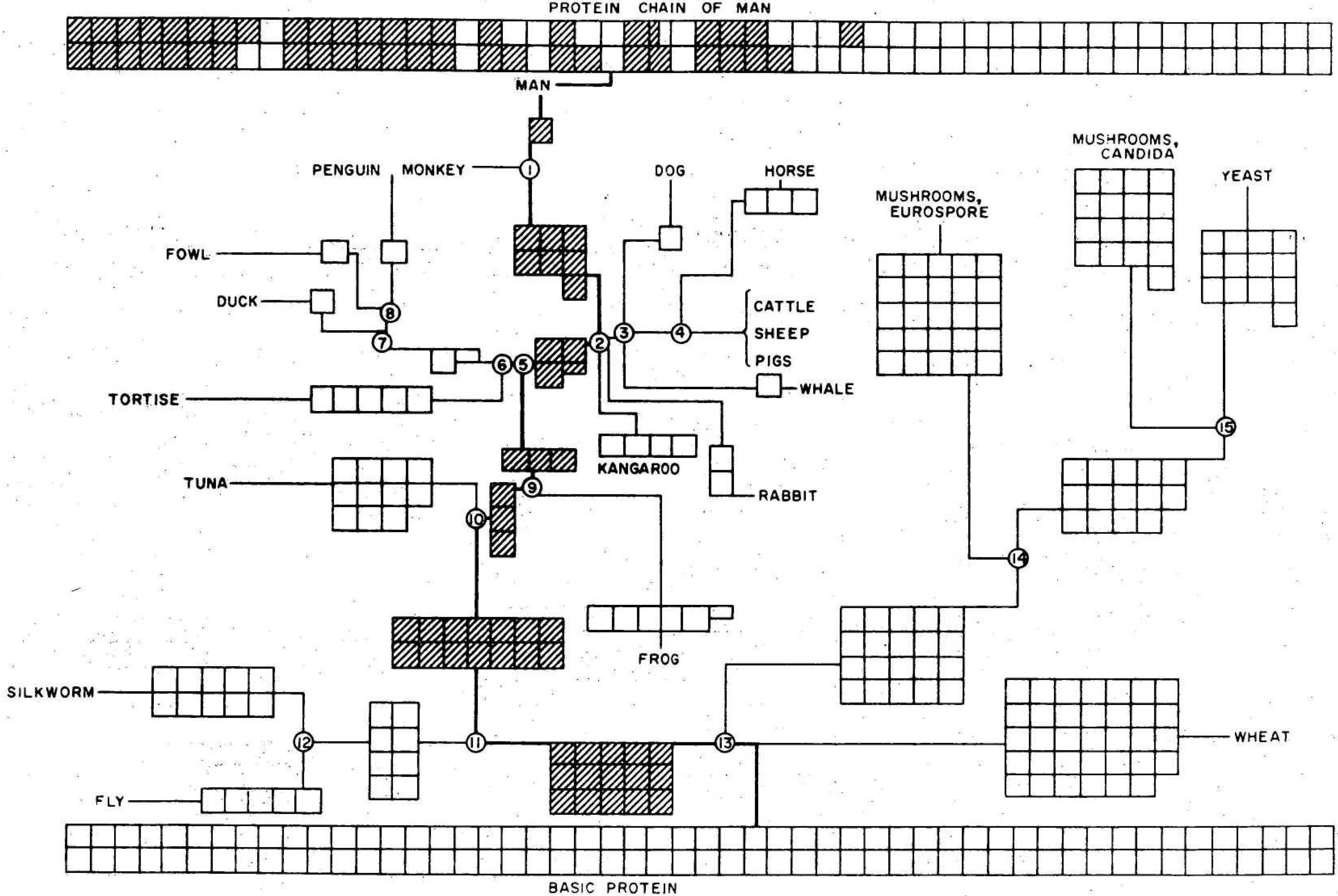


Fig. 30

This is followed by the fifth item, and so on, until a 'tree of approximation' is built.

The third program converts the 'tree of approximation' into a phylogenetic tree. This program examines systematically all the alternatives by cutting ends off branches and 'regrafting' the topologies established in the second program. In this program, millions of configurations are examined, and each is tested by the first program to choose the best. The final result is shown in Figure 30, where on each branch the number of squares indicate the number of mutations which characterize that branch. The shaded squares indicate the protein groups of man.

The close approximation between the final result of the inverted character processing (the phylogenetic tree), and the result of observations through centuries (the conventional 'tree of life') proves the power of this method of establishing 'living' classification systems quickly and automatically. The difference between the two results is only topological; the difference, however, is in favour of the automatic method, as the phylogenetic tree is drawn to scale, while the 'tree of life' is not.

A further example of inverted character processing is described in: 'Point-association of geological data as a quantitative basis for analysis' by Grender, G.C., et al. (1969; published by the Esso Production Research Company, Houston, Texas). Essentially non-quantitative or non-numerical geological parameters such as structure and stratigraphy are treate simultaneously and quantitatively using Kosygin and Voronin's concept of point association of properties. According to this concept, every point in a given space can be thought of as having associated with it any number of properties of any kind, and these properties can be of any quantitative rank from nominal to ratio. Points with given associations can then be counted to produce joint frequency distributions, which characterize the space geologically, and can be used for classification or modelling. An experiment with this idea in the San Joaquin Valley, California, demonstrated not only the feasibility of obtaining such distributions, but also the utility of the system for producing maps and cross-sections of all kinds.

In France, a hydrogeological example is contained in the École Nationale Supérieure des Mines de Paris Record No. LHM/R7014: 'An automatic solution for the inverse problem: Application to the computerisation of the transmissibilities and other aquifer parameters from piezometric data' by Y. Emsellem & G. de Marsily (1970). A method for the automatic determination of trace elements and pollutants is described by Y. Vuillanme in his Ph.D. thesis: 'The utilisation of isotopic and hydrochemical methods in hydrogeology, (Faculté des Sciences de Paris, 1971). The application possibilities of the inverted processing method in the various branches of geology are examined by P. Vases in his Ph.D. thesis: 'Data analysis for the treatment of geological problems' (Faculté des Sciences de Paris, 1964), and by P. Leymarie in his Ph.D. thesis: Automation of graphic constructions (École Nationale Supérieure de Geologie Appliquée, Nancy, 1971). The basic principles of the inverse data processing method are outlined in the book 'ANALYSES DE DONNÉES', published by the University of Paris (1968).

APPENDIX 2: INPUT METHODS AND ORGANIZATION

Punched cards

In many computer installations, the use of the 80-column tab cards is a natural choice, as for the past 80 years the keyboard to punched card equipment has provided the standard means of information transfer in data processing systems. The cards are used not only to enter data into a file, but also to maintain the data in the file.

80-column cards are discrete, and this feature is advantageous in some jobs, where it is convenient to replace an outdated card with a corrected one. For most geological data, however, punched cards are disadvantageous for the following reasons:

i. Since the cards are discrete, they each need some form of identification and sequencing control. If several types of information exist or if data is punched according to various formats, each type must be identified. For example, to record well sample data, on each card, general well data has to be recorded, as well as geographical co-ordinates and identification numbers. This generally means that only about three-quarters of each card is available for real data.

ii. Much geological data is non-data in a sense that, with fixed formats and punched cards, allowances are generally made for the recording of an observation that may be made, but for a variety of reasons has not or cannot be made. Consequently, when data cards are punched, a large amount of available space on each card is left blank.

iii. The translation of the data from the basic documents into punched cards usually involves at least one 'middle-man' (a 'translator', a skilled human interface), but often two or more. Normally, before the data can be captured for data processing, a hard copy document is required which is designed to transact business but not to prepare data for key-punching. Therefore, parts of the original documents are recopied on to coding forms, which are then transmitted to a centralized key-punching section. If the data content is important, it is first key-punched and then key-verified. This laborious process of preparing data for machine use is a time-consuming, tedious, and expensive procedure which introduces continuous opportunities for errors.

iv. Punched cards can be effectively used to deal with data (measurements, results of observations, objective numerals), but they are unsuitable to record information (results of subjective interpretations). Most geological 'data' is 'information', however, which needs to be classified and coded before it can be expressed as data and can be punched on cards. With punched cards therefore, subjectivity enters at two levels in geology: at the time of the interpretation, and also at the time of the codification.

v. When data is contained on punched cards, to rearrange the data, a rearrangement of cards is often necessary. This is achieved by physically sorting and collating them, grouping and regrouping them in various sequences, until the order prescribed in the individual programs is established. In this process, all cards are excessively handled, especially where tabulating equipment is being used, introducing many opportunities to disarrange the decks and thus affect the validity of the final output.

vi. The introduction of punched cards in an organization does not mean a complete changeover from a manual system to an automatic data processing system. It only means that while there was only one system before, now there are two systems to keep up. The manual files cannot be abandoned as a card system always requires a comprehensive supporting system for the constant checks and verifications which are necessary in such an error-prone input system. Furthermore, in designing a format for a card, the tendency is to economize on space, and to record only significant data; but what is not significant today might become important tomorrow: therefore, with a card system a re-entry to the manual supporting system to gather new data is often necessary.

vii. The upkeep of an ADP system built on punched cards (fixed-format systems) is difficult, as from the cards onwards, each data item has to be properly named and classified, before it can become processable, and before it can go into a data file or memory. Especially with field data and geological 'information', the process of giving a definite name to an only broadly defined item not only introduces subjectivity, but also make changes complicated when these are necessary in the light of new evidences, more accurate analyses, or reinterpretations. Each change will have to be followed up by a complete updating process throughout the computer files, memory cells, etc. Since it is a lengthy process and since in most systems the renaming of already processed items cannot be done automatically, these systems are never up-to-date.

Summary. Unless punched cards are automatically produced as a result of completely mechanized processes (x-ray diffractometry, quantometry, automated graphic constructions and analyses (Leymarie, 1971), grainsize analysis, etc.), the input system as a whole can be considered as a total failure, as far as the above stated principles of source data automation are concerned, since there are, at least, three redundant operations before data are in a machine-readable form. The key-punching room is the real bottleneck in most processing operations, yet punched cards are still being used extensively for the input of data and also information in many organizations.

There are probably two main reasons for the still widespread use of punched cards as an input media: piecemeal approaches adopted in many organizations, and the grip of hardware manufacturers on software developments. In both, commercial interests are in the forefront. Keypunch mechanization for the input of information is ideal for suppliers with vested interests in follow-up equipment or systems; it is probably for this single reason that they pour scorn on input solutions which might increase efficiency and self-sufficiency in an organization. A self-contained, easily operable input and processing system would reduce the need for their continuous services, and would reduce dependency. Consequently, neither the suppliers with monopolistic tendencies, nor those who are already committed will readily admit that something is wrong with the universal medicine they are advocating: that punched cards do not necessarily solve all problems and cure all ills, and that their use, in fact, can create new ones. Especially since effective rivals to conventional card-punch input techniques were developed (see following notes), deep evaluations have been made in many enterprises, and criticisms can now be heard loud and clear from those who would like a second chance to start a system anew.

'Perfostyl' prepunched cards

A system using prepunched cards for the input of geological data was developed in the University of Leeds, and was later adopted by the Institute of Geological Sciences, UK, to record borehole data.

In this input system the cards are similar to the conventional punched cards in appearance, but on the 'perfostyl' card (Fig. 31), each character is perforated. A slight pressure with a biro or a pencil at the appropriate point makes a punch when the perforated character falls out, and this punch is recognized by the sensing brushes of normal punched card readers.

Perfostyl cards are an attempt to combine two operations in preparing data for input: in one step, data are formalized and also digitized.

As far as the automatization of source data is concerned, perfostyl cards are an advancement over conventional punched cards. With their use, coding forms and keypunching are eliminated. Decentralization in data preparation is possible, since they can be prepared in the field or in a laboratory. Verification is simple, and it can be carried out by the person who prepares the data; therefore, preparatory errors are minimized. They have one disadvantage, however: they require much card handling and a thorough verification before each use, as, owing to the perforations, unintentional punches o occur occasionally. For this reason, perfostyl cards cannot be considered reliable for continuous use; however, they might be useful for one-step data input procedures, such as library work, borehole data, geochemical results, etc., mainly in piecemeal operations.

Mark-sensing and optical-character reading

In reading a written work or character, the action itself is characterized by the process of perceiving and subsequently recognizing that mark or character. This applies to reading by the human eye as well as to reading by machine. The process by which an optical image is recognized by the human eye had not yet been understood, but the theory exists that the recognition of an object or an image of it is based on earlier perceptions. Therefore, recognizing an image perceived is in fact classifying the thing perceived.

The same applies to machines, but they can only perform such a classification if the features common to a class, that is the class limits, are precisely known. Only after having perceived a character in a certain position, can the machine attempt to classify it.

In the simplest form of mark-sensing the perceiving and the recognizing operations consist only of a test for the presence or absence of a mark in a certain position. For this, simple technical reading devices can be used, which utilize either electrical conductivity (graphic pencil marks), magnetism (special ferrite pencil marks), or photocells (ink marks). Figure 32 shows a general example of a mark-sensing format.

More advanced formats can be achieved, if the marks are applied in the shape of written characters (Fig. 36). Here, man and machine recognize the same thing in different ways: the human eye recognizes the shapes of the written figures, while the machine recognizes the characters by the presence of marks at specified places. This principle has been used in Cambridge, UK, to record geological field data (Fig. 33).

In some related methods, the objective is to give the human observer as clear a view as possible of the subject, or numbers marked as a whole. (Figs 34-40). A marking grid is used in the Netherlands Giro Service (postal, cheque, and clearing service) (see Fig. 36); here, the marking positions are provided with worded digits. Machine recognition is simple, as each digit is represented by a column in which one word has been underlined.

A similar method has also been used at Cambridge University to record geological field data (Fig. 41).

In an advanced form of character-scanning and pattern-recognition (e.g. Litton system) whole pages of data can be converted directly into computer-readable form. Paper is moved at a linear and uniform rate past a row of solid-state self-scanning integrated-circuit photodiode arrays; the recognition is based on a new form of electro-optics. A built-in display unit can provide on-line editing and error-correcting facilities.

Summary. Mark-sensing and optical-character reading devices fall into the same category as the perfostyl cards. They allow simultaneous formalization and digitization, and the possibility of decentralization in data preparation on the one hand, but they are fragile and susceptible to unintentional markings (dirt) on the other. There is one additional disadvantage, however: mark-sensing and character-sensing formats need special reading equipment at the information processing stage, and these are increasing the cost of interfacing the input method with the computer. Experiments carried out at the University of Cambridge proved, however, that they can be useful in special piecemeal operations (such as logging cliff faces while hanging onto a climbing rope in sub-zero temperatures in Iceland) where considerations in simple input are overriding those of economy in subsequent processing (Cutbill*, 1971).
BMR File 1972/156 details a proposal of using mark-sensing forms and an

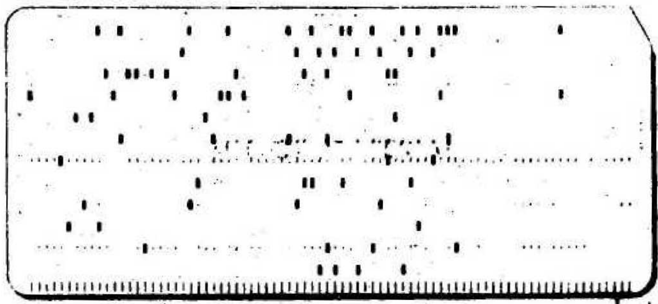
---

*CUTBILL, J.L., 1971 - DATA PROCESSING IN BIOLOGY AND GEOLOGY, London, Academic Press.

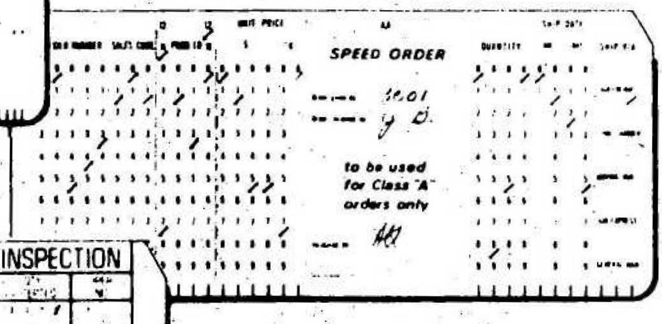
12 ROCK CARD 2 11 PROJECT / R /			INSTITUTE OF GEOLOGICAL SCIENCES MINERALISATION				Borehole number	BORE HOLE DEPTH (Feet)	
								Minimum	Maximum
0 0	0	0 0 0 0	0 0 0	0 0 0	0 0 0	0 0 0 0 0 0 0 0 0 0 0 0	0 0	0 0 0 0 0	0 0 0 0 0
2 4	6	8 10 12 14	16 18 20	22 24 26	28 30 32	34 36 38 40 42 44 46 48 50 52 54 56	58 60	62 64 66 68 70	72 74 76 78 80
I I	I	I I I I	I I I	I I I	I I I	I I I I I I I I I I I I	I I	I I I I I	I I I I I
2 2	2	2 2 2 2	2 2 2	2 2 2	2 2 2	2 2 2 2 2 2 2 2 2 2 2 2	2 2	2 2 2 2 2	2 2 2 2 2
3 3		3 3 3 3	3 3 3	3 3 3	3 3 3	3 3 3 3 3 3 3 3 3 3 3 3	3 3	3 3 3 3 3	3 3 3 3 3
4 4	4	4 4 4 4	4 4 4	4 4 4	4 4 4	4 4 4 4 4 4 4 4 4 4 4 4	4 4	4 4 4 4 4	4 4 4 4 4
5 5	5	5 5 5 5	5 5 5	5 5 5	5 5 5	5 5 5 5 5 5 5 5 5 5 5 5	5 5	5 5 5 5 5	5 5 5 5 5
6 6	6	6 6 6 6	6 6 6	6 6 6	6 6 6	6 6 6 6 6 6 6 6 6 6 6 6	6 6	6 6 6 6 6	6 6 6 6 6
7 7	7	7 7 7 7	7 7 7	7 7 7	7 7 7	7 7 7 7 7 7 7 7 7 7 7 7	7 7	7 7 7 7 7	7 7 7 7 7
8 8	8	8 8 8 8	8 8 8	8 8 8	8 8 8	8 8 8 8 8 8 8 8 8 8 8 8	8 8	8 8 8 8 8	8 8 8 8 8
9 9	9	9 9 9 9	9 9 9	9 9 9	9 9 9	9 9 9 9 9 9 9 9 9 9 9 9	9 9	9 9 9 9 9	9 9 9 9 9
2 4	6	8 10 12 14	16 18 20	22 24 26	28 30 32	34 36 38 40 42 44 46 48 50 52 54 56	58 60	62 64 66 68 70	72 74 76 78 80

XAUS-2-170

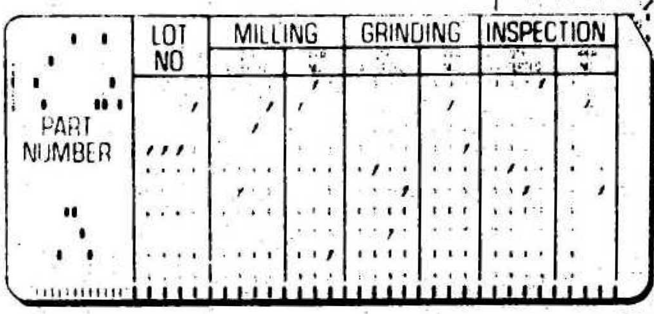
Fig. 31



PUNCHED CARDS



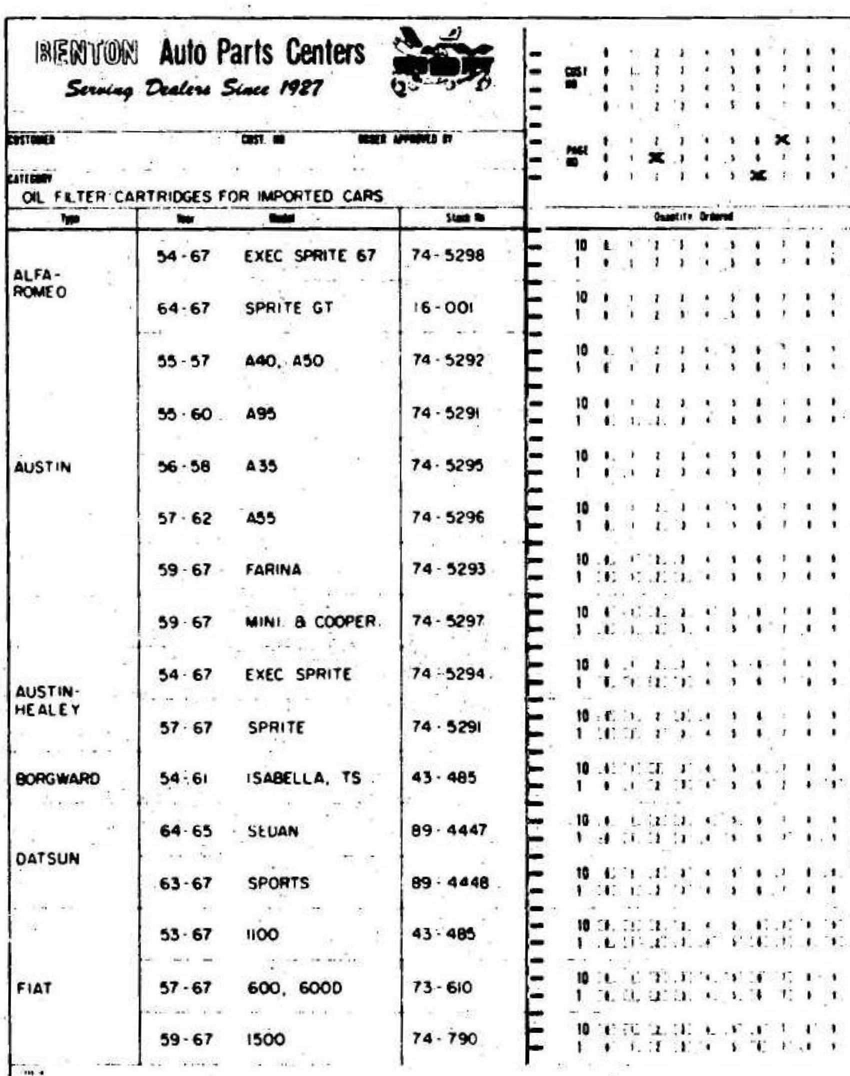
PENCIL-MARKED CARDS



PENCIL-MARKED AND PUNCHED COMBINED

With the Motorola MDR, you are free to design each document for optimum human intelligibility and accuracy. The MDR will handle standard punched cards or documents up to $8^1/_2$" wide by any length. The cards can contain information that has been pencil marked, punched or both on one card. Format of the marks and punching conforms to normal Hollerith coding with a maximum of 12 data positions per column. Formats are identical for documents and cards. This is made possible by holding the width of the reading area to a strip $3^1/_4$" wide — the industry standard for tab cards.

The $3^1/_4$" strip runs the length of the card or page and is positioned along the right-hand edge of the document. The remainder of the document may be used for whatever purpose the user's application calls for.



BENTON Auto Parts Centers

Serving Dealers Since 1927

CUSTOMER — CUST. NO — ORDER APPROVED BY

CATEGORY: OIL FILTER CARTRIDGES FOR IMPORTED CARS

Type	Year	Model	Stock No
ALFA-ROMEO	54-67	EXEC SPRITE 67	74-5298
	64-67	SPRITE GT	16-001
AUSTIN	55-57	A40, A50	74-5292
	55-60	A95	74-5291
	56-58	A35	74-5295
	57-62	A55	74-5296
	59-67	FARINA	74-5293
	59-67	MINI & COOPER	74-5297
AUSTIN-HEALEY	54-67	EXEC SPRITE	74-5294
	57-67	SPRITE	74-5291
BORGWARD	54-61	ISABELLA, TS	43-485
DATSUN	64-65	SEDAN	89-4447
	63-67	SPORTS	89-4448
FIAT	53-67	1100	43-485
	57-67	600, 600D	73-610
	59-67	1500	74-790

Record 1975/9

FIG 32

176

GEOLOGICAL FIELD DATA SHEET, CAMBRIDGE

STATION NUMBER

A	B	C	D	E	F	G	H	J	K
000	100	200	300	400	500	600	700	800	900
00	10	20	30	40	50	60	70	80	90
0	1	2	3	4	5	6	7	8	9

SHEET NUMBER

00	10	20	30	40	50	60	70	80	90
0	1	2	3	4	5	6	7	8	9

WRITE STATION AND SHEET NUMBERS ON STUB

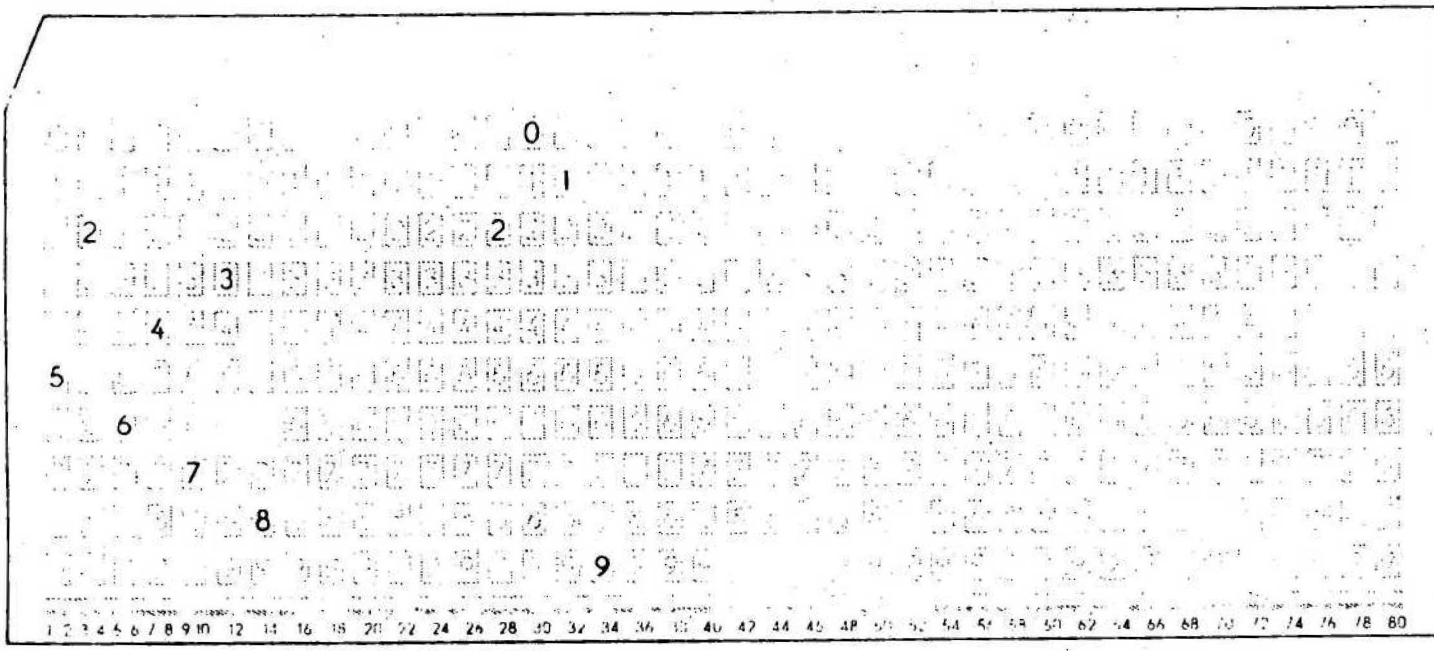
0	1	2	3	4	THICK	5	6	7	8	9
0	1	2	3	4	(M)	5	6	7	8	9
00	05	EXPOSED	COVERED			RED	PRD	GN	GY	OTHER
F SL	M SL	CSL	VFSS	FSS		MSS	CSS	VCSS	CON	PEBBLY
GRAD	SHP	SMTH	SCR	TOOL	MKS	TRFOS		MDCR		RIPPLE

NO	CEM	$CaCO_3$	ROCK	DOL	FLAT	SYP	ASR	PLX	TRX
LIN	DEF	CONCR - CO_3 OR STUB			FOSSILS	NO	VERT	OTHR	TR
NO DIR	2 DIR	1 DIR				000	100	200	300
00	10	20	30	40	50	60	70	80	90
STUB ← SPEC				RESTART	1	2	SPARE	4	CANCL

Record 1975/9

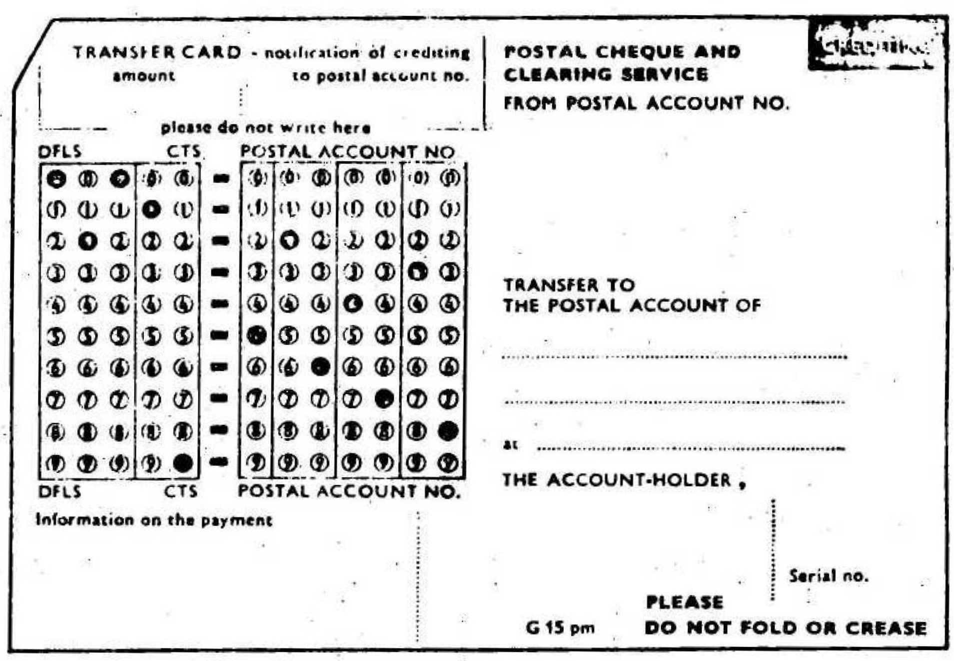
Fig. 33

XAUS-2-169



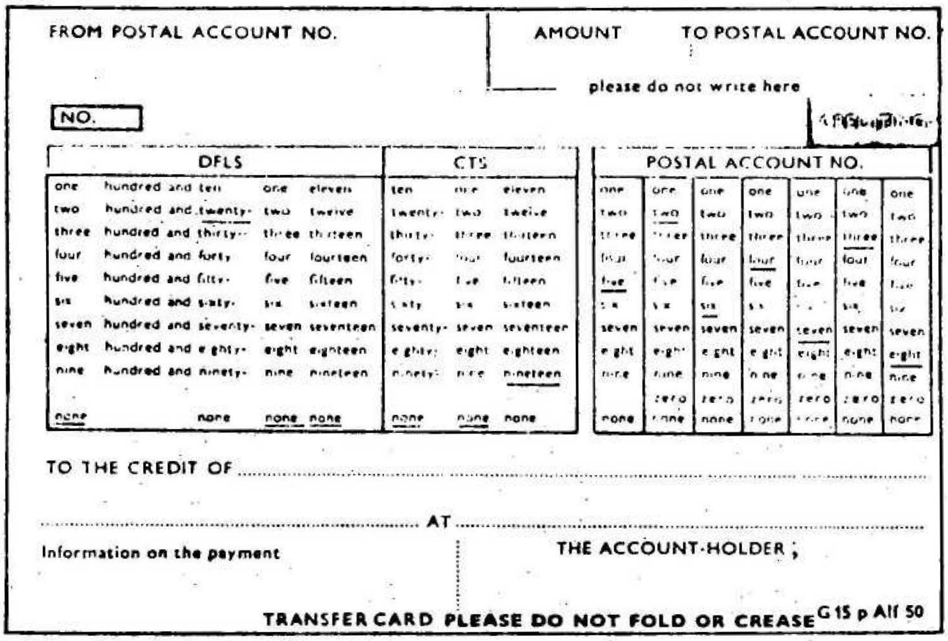
0
1
2
2
3
4
5
6
7
8
9
1 2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80

Fig. 34



TRANSFER CARD - notification of crediting
amount
to postal account no.
please do not write here
DFLS
CTS
POSTAL ACCOUNT NO
DFLS
CTS
POSTAL ACCOUNT NO.
Information on the payment
POSTAL CHEQUE AND
CLEARING SERVICE
FROM POSTAL ACCOUNT NO.
TRANSFER TO
THE POSTAL ACCOUNT OF
at
THE ACCOUNT-HOLDER ,
Serial no.
PLEASE
DO NOT FOLD OR CREASE
G 15 pm

Fig. 35



FROM POSTAL ACCOUNT NO.

AMOUNT

TO POSTAL ACCOUNT NO.

please do not write here

NO.

DFLS

one	hundred and	ten	one	eleven
two	hundred and	twenty-	two	twelve
three	hundred and	thirty-	three	thirteen
four	hundred and	forty-	four	fourteen
five	hundred and	fifty-	five	fifteen
six	hundred and	sixty-	six	sixteen
seven	hundred and	seventy-	seven	seventeen
eight	hundred and	eighty-	eight	eighteen
nine	hundred and	ninety-	nine	nineteen
none		none	none	none

CTS

ten	one	eleven
twenty-	two	twelve
thirty-	three	thirteen
forty-	four	fourteen
fifty-	five	fifteen
sixty-	six	sixteen
seventy-	seven	seventeen
eighty-	eight	eighteen
ninety-	nine	nineteen
none	none	none

POSTAL ACCOUNT NO.

one	one	one	one	one	one	one
two	two	two	two	two	two	two
three	three	three	three	three	three	three
four	four	four	four	four	four	four
five	five	five	five	five	five	five
six	six	six	six	six	six	six
seven	seven	seven	seven	seven	seven	seven
eight	eight	eight	eight	eight	eight	eight
nine	nine	nine	nine	nine	nine	nine
	zero	zero	zero	zero	zero	zero
none	none	none	none	none	none	none

TO THE CREDIT OF ....

.... AT ....

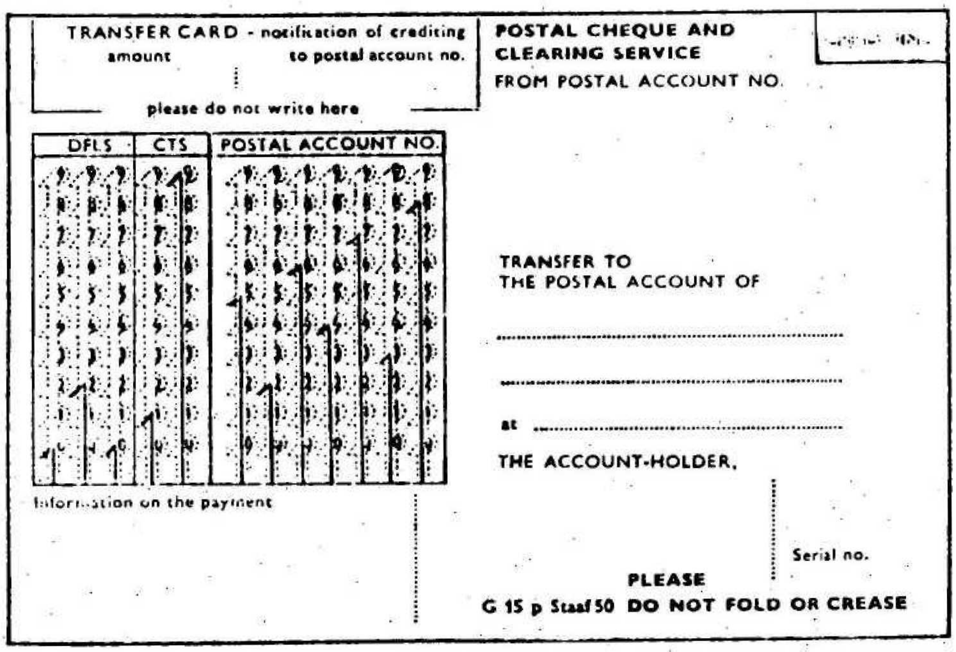
Information on the payment

THE ACCOUNT-HOLDER ;

TRANSFER CARD PLEASE DO NOT FOLD OR CREASE G 15 p Alf 50

Owing to the construction of numerals in languages such as Dutch, Danish and German the total number of columns in the alphabetic marking-grid can be reduced by one; in this manner an arrangement of columns is obtained which corresponds to the pronunciation of those numerals.

Fig. 36



TRANSFER CARD - notification of crediting

amount

to postal account no.

please do not write here

DFLS | CTS | POSTAL ACCOUNT NO.

Information on the payment

POSTAL CHEQUE AND CLEARING SERVICE

FROM POSTAL ACCOUNT NO.

TRANSFER TO

THE POSTAL ACCOUNT OF

....

....

at ....

THE ACCOUNT-HOLDER,

Serial no.

PLEASE

G 15 p Staaf 50 DO NOT FOLD OR CREASE

Fig. 37

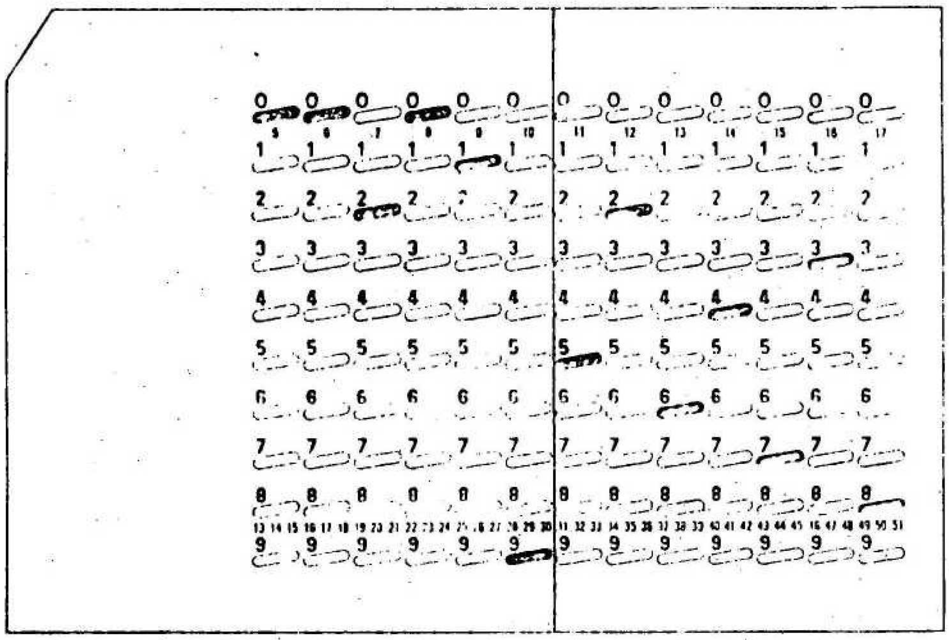


Fig. 38

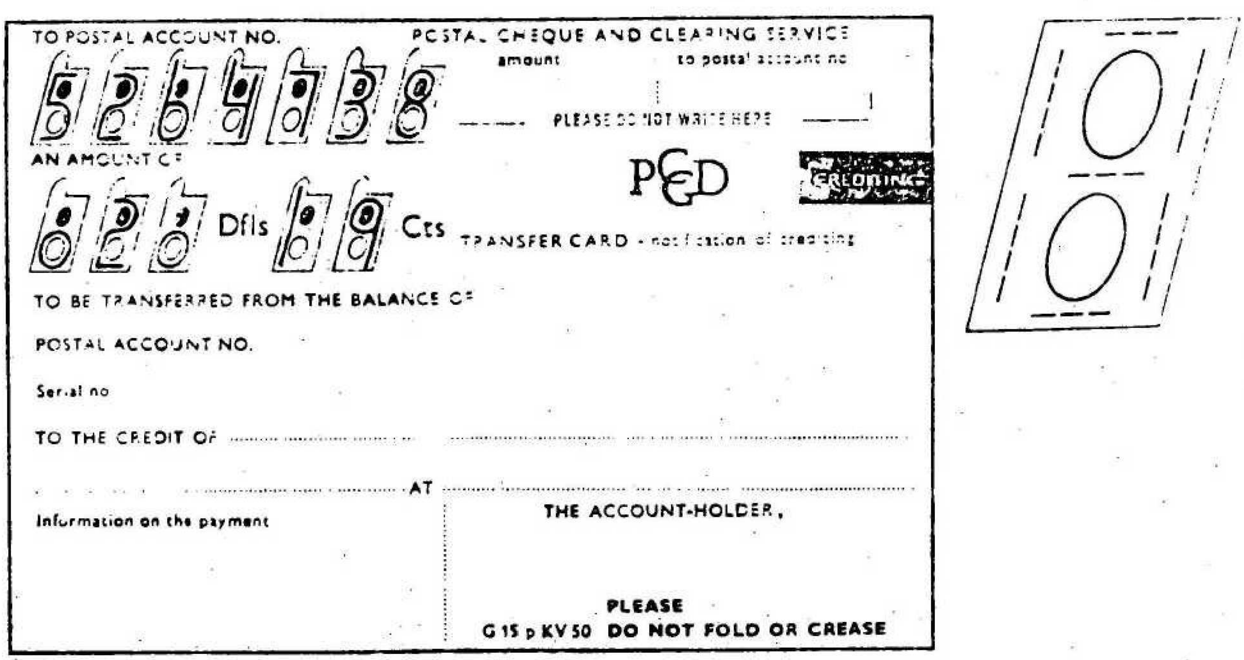

TO POSTAL ACCOUNT NO.
POSTAL CHEQUE AND CLEARING SERVICE
amount
to postal account no
PLEASE DO NOT WRITE HERE
AN AMOUNT OF
PCD
Dfls
Cts
TRANSFER CARD - notification of crediting
TO BE TRANSFERRED FROM THE BALANCE OF
POSTAL ACCOUNT NO.
Serial no
TO THE CREDIT OF
AT
Information on the payment
THE ACCOUNT-HOLDER,
PLEASE
G 15 p KV 50 DO NOT FOLD OR CREASE

Fig. 39

please do not write here

POSTAL CHEQUE AND CLEARING SERVICE

TRANSFER CARD - notification of crediting

FIRST DIGIT	SECOND DIGIT	THIRD DIGIT	FOURTH DIGIT	FIFTH DIGIT	SIXTH DIGIT	SEVENTH DIGIT
5	2	6	4	7	3	8

TO THE ABOVEMENTIONED POSTAL ACCOUNT AN AMOUNT OF

GUILDERS IN HUNDREDS	GUILDERS IN TENS	GUILDERS		TEN CENT UNITS	CENTS	
0	2	0	Dfls	1	9	Cts

TO BE TRANSFERRED FROM THE BALANCE OF

PCGD

POSTAL ACCOUNT NO.

Serial no.

TO THE CREDIT OF ....

.... at ....

Information on the payment

THE ACCOUNT-HOLDER.

PLEASE

G 15 p Hor 50 **DO NOT FOLD OR CREASE**

1/A 1/B 2 3 4

✲ 1. Letter / No/ Unit /Sequence/Control Scientific

2C. Origin/Direction /Method/Inclination / C = Uphill, X = Downhill
(=Taping, C = Aneroid, X = Vert. Int., S = Paced, Z = Direct Measurement)

2S. Base/Thickness/Taped, Estimated/Exposure %/Covered, exposed, float

3C. Measurement Unit /Tape Length

3S. Dip Direction/ Dip Amount / C= Normal, X = Overturned / Photo

4. Stub/Specimen No/Specimen No/Photo No

5. Bed Type Relations/%In Unit/Colour/Saccharoidal, Cruddy

6. Bedding/Lith/ Por/C= Nodules/Grain Size /Sort/Deform = X

7. Thickness of Bed Types - Mean/Min/Max/Erosion Resistance 0-9

BED TYPE RELATIONS
- 0 = One Bed Type In Unit
- 1 = Random Alternation
- 2 = Repetitive Sequence
- 3 = Simple Sequence
- 4 = Gradation

COMPOSITION
- X = Mineral
- S = Sedimentary
- T = Igneous
- Z = Metamorphic
- C = Undifferentiated

GRAIN SIZE
- 0 = Clay
- 1 = Silt
- 2 = Fine Sand
- 3 = Medium Sand
- 4 = Coarse Sand
- 5 = Gravel
- 6 = Pebble
- 7 = Cobble
- 8 = Boulder

ABUNDANCE
%0 - 99, X, Mid-Class

POROSITY
0 - 3

STRATIFICATION
- 0 = < 2 cm.
- 1 = 2 - 1
- 2 = 1 - 5
- 3 = 5 - 60
- 4 = 60 - 120
- 5 = > 120
- C = Contorted
- S = Planar
- X = Small X-Bed
- Z = Large X-Bed
- T = Structureless

ROCKS
- 00 - Undifferentiated
- 23 - Amphibolite/Basic
- 44 - Feldspathite/Acid
- 77 - Quartzite
- 22 - Psammite
- 33 - Pelite
- 66 - Carbonate (Undif)
- 11 - Limestone
- 99 - Dolomite
- 55 - Carbonaceous
- 88 - Ironstone
- 40 - Evaporite
- 12, 57, 89, 48, 94 Spare

STUB S = Notes On Margin

COLOUR
- 0 - Yellow
- 1 - Grey
- 2 - Green
- 3 - Brown
- 4 - Black
- 5 - Orange
- 6 - Purple
- 7 - Red
- 8 - Blue
- 9 - White

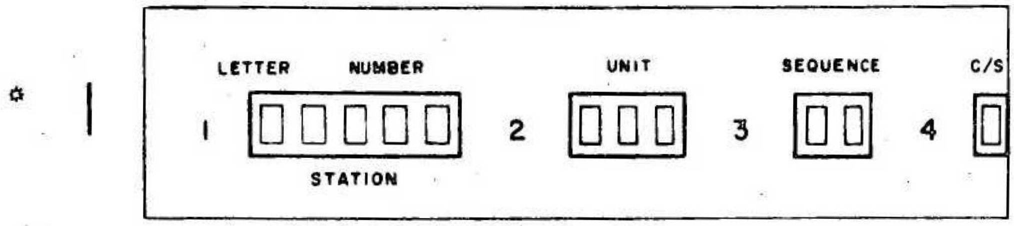
MINERALS
- 00 - Undifferentiated
- 15 - chert
- 29 - Quartz
- 54 - Feldspar
- 82 - Sheet Silicate
- 93 - Ferromag
- 65 - Glauconite
- 61 - Pyrite
- 31 - Phosphate
- 71 - Calc Carb
- 76 - Dolomite

FORM
- 0 - Unspec
- 1 - Whole Rock
- 2 - Clast
- 3 - Fossil
- 4 - Pellet/Lump
- 5 - Oolite
- 6 - Intraclast
- 7 - Nodule/Concretion
- 17 - In Situ
- 16 - Pseudo
- 32 - Round
- 67 - Angular
- 28 - Platy
- 75 - Banded
- 84 - Spotted
- 45 - Irregular
- 64 - Intrusion
- 39 - Lava
- 51 - Pyroclastic
- 73 - Fissile
- 92 - Schistose
- 49 - Folded
- 53 - Ruptured
- 21 - Well Jointed

NOTE
1 PRINT NEATLY
2 WRITE LARGE
3 CLOSE LOOPS
4 JOIN LINES
5 SIMPLE SHAPES

0 1 2 3 4 5 6 7 8 9

C S T X Z



Record 1975/9 XAUS--2 -172

Fig. 41 Main sensing format code used by J.L.Cutbill & D Piper Sedgwick Museum, Cambridge.

optical reader to process well and sample data in the Core and Cuttings Laboratory.

Keyboard to computer input methods

For recording geological data, keyboard-to-computer input methods are being used extensively both in the US and in Canada (Smithsonian Institution, Canadian Geological Survey, major oil companies, etc). There are several types of connection between a keyboard and a 'stand-alone' terminal; between a 'remote terminal' and a computer; and between a small computer and a 'central processor':

'off-line'. A direct entry is used in off-line arrangements, where the input informations are immediately treated by a processor or a computer in lots ('batch-processing'). A 'remote batch terminal' refers to a 'stand-alone' terminal at a distance from a central processor.

'on-line'. In these systems, the input of information is controlled by a 'relay multiplexer', which directs the lines arriving from the different sources into a sequence, assembles the messages, controls validity, etc. A facility exists here to enter and transmit data from point A to point B, but a 'conversation' is not implied.

'multiprogrammation'. This is a technique employed in the more complex systems to utilize the central memory power to the maximum. Several programs can use the well addressed, determined, and individually 'named' elements of the central memory simultaneously, and the whole process is controlled by a 'monitor'.

In all these keyboard-to-computer input methods, a direct or indirect contact is implied between the data processing terminal and the computer, providing sufficient power only at the required times (thus reducing hire charges), while ensuring a speedy response whenever needed. In these methods, retranscriptions and key-verifying of punched data are removed from the data preparation cycle, and, therefore, for source data automation, these methods are generally more attractive, than punched cards. Remote data processing terminals coupled to computers directly or indirectly were found to be both operationally and economically reasonable in those fields of activity where the speed of response is essential, especially in military applications, airline reservations, banking, brokerage and stock quotations, credit and real-estate checkings, hospital administration and medical data processing, manufacturing and merchandising inventories, etc.

However, simple input organization (and, therefore, the least data preparation), efficient utilization of available power (modular structure and multiprogrammation), and flexibility to adopt published programs written in various languages (automatic interfacing) would be more important in geology than speedy response. Therefore, since they are oriented towards achieving a goal which is not really essential in geology, keyboard-to-computer direct input methods fall short of an ideal for the following two main reasons:

i) all data to be encoded should be carefully prepared and every item and field precisely defined before data reaches the keyboard. The difficulties in preparing geological data for fixed-format processing have already been outlined in this Record.

ii) it is difficult for the operators to correct errors on a segmental access medium such as magnetic tape encoder, unless the keyboard is connected with an expensive cathode-ray display tube (as in Sanders, Viatron, etc.)

Recent 'stand-alone' CRT model terminals are capable of storing complete forms (titles, headings, and input data formats) on a magnetic tape, and displaying them on the screen one or more at a time (in the Sanders medical data processing system, for example, over one thousand characters can be shown on the screen at any one time). Data input is by a keyboard; each input item is immediately displayed on the screen under or against the title according to the format. Input errors are eliminated, as (1) the machine does not accept improper format, and (2) the operator can visually verify the correctness of each item which has been keyed in. When both the machine and the operator accept what is on the screen, titles, formats and data are stored together on a second magnetic tape, which is forwarded to the computer (for further details, see BMR File 70/986, parts 1 and 2). This arrangement could possibly replace the Friden typewriter as an input instrument for the semantic system.

Natural-word input systems

In these systems the natural language (words and word-roots) is processed. The individual words are picked out and recognized when (1) the title of the publication, (2) the abstract, or (3) the whole text is being fed into the system.

The Institut Francais du Petrole (IFP) is currently experimenting with such a system. In their natural-word system, only the words included in the title are processed, but not yet the whole abstract, or text. A retrieval

1238057224 UAKKAU

- CONTENTS OF SOME TRACE COMPONENTS IN TADZHIK DEPRESSION PETROLEUMS.
- ZUL FUGARLY DI AZIZOV NA NUMANOV IU
- (USSR).
- UCH. ZAP. AZERB. GOS. UNIV., SER. KHIM. NAUK NO. (),25-32(1968)

COPPER PETROLEUM. TRACE ELEMENTS PETROLEUM. PETROLEUM TRACE ELEMENTS. ELEMENTS TRACE PETROLEUM. VANA DIUM PETROLEUM. NICKEL PETROLEUM

1263757224 SVGLAZ

- STRONTIUM IN FORMATION WATERS OF THE EASTERN CARPATHIAN SYNCLINE AND ITS SIGNIFICANCE DURING PROSPEC
- TING FOR OIL AND GAS.
- SHCHEPAK VM MIGOVICH VI
- (INST. GEOL. GEOKHIM. GORYUCH. ISKOP., LVOV, USSR).
- SOV. GEOL. 13(2),7-9(1970)

GASES OILS PROSPECTING. STRONTIUM FORMATION WATERS. PROSPECTING OILS GASES. OILS GASES PROSPECTING. FORMATION WATERS SR. WATERS FORMATION SR

Fig 42

example is shown in Figure 42. The search program included the following commands:

Selection 1: ' - each document which contains one of the following terms:

PETROLEUM ....

OIL .... '

Selection 2: ' - each document which contains one of the following terms:

TRACE ELEMENT

NICKEL

VANADIUM

TITANIUM

CHROMIUM

STRONTIUM'

Selection 3: ' - each document which <u>does not</u> contain one of the following terms:

..... DESULFURISATION

DEMETAIL .....

CONTAMIN .....

CATALY .....'

An important aspect of a document retrieval system is to retrieve all documents which contain relevant information. In the IFP system, this is achieved by positive and negative selection commands, each of which refers to individual words. In more advanced natural-word systems, however, the relevance of retrieval is based on information content.

In the natural-word system used by the Subsurface Laboratory, University of Michigan, for example, the natural-word descriptors found in a text are structured on levels, which are defined as derivatives of concepts.

The technique of co-ordinate reduction is used for the automation of the text analysis, and for the selection of a small group of descriptors. The selected group of descriptors is arranged in clusters according to their associations, and their degree of relevance to a concept. In this system, therefore, the relevance of the retrieved references is determined by the search definition, which is based on descriptors and concepts - not words.

According to the IFP experiences, the establishment of a natural-word system is difficult owing to synonymity, inconsistent use of words, superfluous vocabularies in texts, divergency and incongruency between the various literary styles, etc. A natural-word system is not yet considered to be a practical proposition in France.

Free-format input systems

In these systems keywords are used for the automation of source data. However, the keywords are somewhat special: they are equivalent (in purpose rather than form) to the headings used on the coding forms for punched cards. In a fixed format geochemical data processing system, for example, the percentages of the various elements determined in a sample are punched in a certain order of the cards. Using a positional code (format statement), the input program recognizes the order, pairs the numbers (results) to the headings (element or mineral names), and determines that the numerals are expressing percentages. In a free-format input system copper (or Cu) and percentage (or %) are keywords; it is sufficient, therefore, to punch in any position:

/Cu = 79%/

Such an entry in a free-format input system will put the main titles (sample number, location, etc.) and all information components between the slashes (headings and results) in a 'data state' (individually addressed and paired together in memory). The recently (1974) introduced 'INFOL' system of CSIRO/BMR falls into the same category.

The absence of involved codification and the individual recognition of each input item are the advantages of this system. In the Ecole Nationale Superieure des Mines de Paris, worldwide inventories of minerals and usable substances, catalogues of key-maps showing mine positions, etc. are being created using this system (see Fig. 43 for an output example; the selection for obtaining this result was carried out in a manner similar to that of the natural-word example). In the Centre dde Recherches Petrographiques et Geochimiques in Nancy, the input of geochemical data is carried out this way.

The usefulness of the free-format input method has already been proved in commercial and routine type of piecemeal operations, and - owing

CARTES D'AUSTRALIE SELECTIONNEES
****

****

```
NO. DE L'ANALYSE=C0005
DOCUMENT=C
BIBLIOTHEQUE=P LAFFITTE
TITRE=ATLAS OF AUSTRALIAN RESOURCES
FEUILLE=MINERAL DEPOSITS
ECHELLE=6000000
EDITION=1952
PROJECTION=POLYCONIQUE
LIMITE NORD=S9D30M
LIMITE SUD=S44D
LIMITE OUEST=E113D15M
LIMITE EST=E153D34M
PAYS=AUSTRALIE
ANALYSEUR=LESAGE
CONTENU PRINCIPAL DE LA CARTE=MINE
RESEAU=GEOGRA
FOND GEOGRAPHIQUE=ALT
                  CRS D'EAU
                  MOY. NOMS
                  CH FER
                  LIM ADM
CARTE MINIERE ET METALLOGENIQUE=LOCAL
                                TAILLE
                                METAL
                                NON METAL
                                CONSTRUCT
CARTE DES DEPOTS HOUILLERS=LOCAL
                           TAILLE
                           CHARBON
AL=BAUXITE
SB=+
SN=+
AS=+
ZR,(HF)=ZIRCON
H=+
U,(RA,ACT.,TRANSU.)=+
V=+
ZN=+
TA,(NB)=COLUMBITE
        TANTALITE
FE=ILMENITE
   JAROSITE
BE=BERYL
BI=BISMUTH
CD=+
CR=+
CO=+
PB=+
TI=RUTILE
LI=+
MATIERES PREMIERES=ARGILE REFRACTAIRE
                   DIATOMITE
                   DOLOMIE
                   EMERI
                   GLAUCONIE
                   GRAPHITE
                   CALCAIRE
                   MICA
```

to its retrospective search facility - also in research type of studies. One drawback prevents its adoption as a national data processing system: a synonym-free thesaurus of 'keywords' which is acceptable, understandable, enforceable, and compatible with all others on a national scale would still have to be created. The present aim in France is to extend the method, and use it, whenever it is advantageous, parallel with semantics and other systems on a national scale (see Fig. 1.)

Input organization

Once it is decided to use a computer in the development process for a geological data and information storage and retrieval system, the next step is to decide where and how to begin recording data in a machine-readable form as close to the source as possible, and thereafter process it with as little human intervention as possible.

There is a spectrum of methods to build systems so that they can be tailored for a specific set of requirements; they range from applications in which the user has no alternatives to methods in which he has complete freedom. In the most completely specified systems the user can only enter data values. This approach tends to be satisfactory only when the problem is relatively small and well defined. Some systems allow more freedom through the use of parameter values as well as data values. In other systems, a number of options are provided, while a more general approach is represented by user oriented languages both for input and output. These give the user a relatively flexible method of specifying his problem, but require less effort than would be required to write a program in general-purpose language. In all geological data processing systems, however, the following phases need to be distinguished:

1. An initial phase, when the objectives of the system are defined; when those stages of the work are decided upon where the automatic processing of data is necessary; and when the feasibility of using a computer both economically and conceptually is established.

2. A second phase, when those data items and informations are determined which will be treated by the machine.

3. A study phase, to investigate the possibilities of utilizing known connections between data items and between groups of data, and to establish a general structure for data treatment.

4. During formalization, the man-machine, machine-to-machine, and machine-to-man interfaces are established, and methods chosen for input-output, codification, indexing, etc.

5. Data preparation stages, when the existing data systems and procedures are reorganized according to the requirements of the agreed process of computerization.

6. A system preparation phase, for the entering of data documents into computer memory (preparation of data files, magnetic tapes, etc). In this phase, programs are written for the storage and treatment of data.

7. Machine treatment, error, detection, controls, corrections.

8. Study of results, provisional conclusions, modifications to treatment.

9. Readjustment phase: new treatments, controls, corrections.

10. Final conclusions (or a return to 8, or to 6 for more data; or a return to 1, for a reappraisal of the whole system).

In all these phases several people play important roles. Top management defines objectives and decides on entry to computerization. Geologists and geophysicists define the problems which need to be studied; they gather and prepare the data required for the studies, and they interpret and criticise the results. Computer scientists (programmers, systems analysts) are occupied with the machine (input, output, listing, sorting, machine treatment, etc.). However, between these two main spheres of activity there are overlapping interest fields, and these are covered in France by 'information geologists', who are trained both in geology and in computer work.

Information Science in France is considered to cover both theoretical and practical aspects of working with information. In the United States the accent is on the theoretical aspect almost exclusively, while in Great Britain information science as such is equated to only one practical application: librarianship (Hanson, 1971). By contrast, 'Informatique Geologique' is a recognized profession in France: it is a speciality. Persons occupying these positions might be geologists with a strong leaning towards computers, or programmers with a diploma in geology. Information geologists are there, in each organization, to carry out pilot work and critical studies concerning the suitability of certain treatments for a given set of data, to attempt to automate some data, and to form a liaison between those who write programs for computers and those who utilize the programs. At the input stage, they are the ones who formalize data, while after the treatment, they prepare the results in such a form that geologists can interpret them.

The information departments which handle all technical information as well as documentations are considered to be 'nerve centres' in such national geological organizations as BRGM, IFP, SNPA, etc. In their geological data processing systems, the biggest problems are unloaded onto the information geologists, whose task is to formalize raw data. They are the ones who bridge the gap between geological information on the one hand, and machine processable data on the other.