Computers & Geosciences, Vol. 2, pp. 347–349. Pergamon Press, 1976. Printed in Great Britain

THE GEOLOGICAL COMPUTER[1]

KEITH G. JEFFERY
Computer Unit, Institute of Geological Sciences, Exhibition Road, London SW7 2DE, England

and

ELIZABETH M. GILL
Atlas Computing Division, Rutherford Laboratory, Chilton, Didcot OX11 0QY, Oxon, England

(*Received* 27 *November* 1975)

Abstract—A "machine" is proposed which works with a geologist (or other scientist) to help him in his everyday work. The machine consists of a relational data base, highly structured processing software designed to reflect geological thinking and input/output interface software to aid a user of the system. The G-EXEC System represents a partial implementation of such a "geological computer".

Key Words: Data files, Data systems, G-EXEC, User interface, General.

INTRODUCTION

It is strange that the adjective "geological" has not been applied to the computer. Geologists, in the past, have taken a general-purpose implement and have modified it for their purpose, adding the adjective geological to indicate this change. Two well-known examples are the geological hammer and the geological map; in both instances the general concept has been modified to suit the geological purpose. The computer is another general-purpose implement which, with suitable modification, should suit the requirements of the geologist. The logical design of such an implement now is considered.

There is a classical sequence of events in the design of a system to suit a requirement. This sequence is:

(a) decide required output(s);
(b) consider available input(s);
(c) design a "machine" to produce (a) from (b).

In the situation of the geological computer the outputs and inputs span a great diversity both in content (that is subject-area within the science) and in form (for example tables of results, maps, diagrams). Any machine designed to handle the task must be capable of processing data of different types through a variety of logical processes or subtasks. So far, a simple general scientifically orientated machine has been defined. The geological aspect consists of the accumulated knowledge collected by the geologists themselves (in the form of a data base representing "geological memory"), and knowledge about the techniques of manipulating these data to obtain results meaningful to a geologist (in the form of software written by or for geologists). The "geological computer" is a machine which extends the data accessible to the individual geologist by adding to it the accumulated data of the other geologists involved in the system. Its characteristics include the ability to respond rapidly to his requirements for the information which is relevant to the current problem or task, and to supply that information in a form which is readily usable by the geologist.

Such a machine has four components;

(a) a user interface for input;
(b) a user interface for output;
(c) geologically orientated software for processing;
(d) a representation of the world as collected by geologists in a data base for memory.

A partial implementation of this machine has been attempted as the G-EXEC system (Jeffery and Gill, 1976).

THE USER-INTERFACE FOR INPUT

The inputs may be classified (in systems terms) into commands, data description, and data. The commands have the purpose of communicating to the system the user requirement of the system. Their function is to activate units of software, and to activate data sets.

The data description describes the data set to both the software and to the user—it forms a communication medium between the data and the two processing machines to act upon it, namely the software in the computer and the mind of the geologist (or other user). The data description must describe completely the content of the data (that is the significance of the data items in a domain) and the structure of the data (that is the interrelationships between data items and sets of data items).

The data themselves are descriptions of items in the real world and, as such, must convey information about the item as precisely as possible to the user.

THE USER INTERFACE FOR OUTPUT

This interface must provide the information required by the scientist in an acceptable form. From the system design point of view, the required output modes may be classified as direct output, edited output, and special output.

Direct output is print or graphic presentation of

[1]This paper is published with permission of the Director, Institute of Geological Sciences, and of the Director, Atlas Computing Division, Rutherford Laboratory.

information in general style. Examples are tables of results, standard X–Y graphs, maps of point locations, and results of standard statistical analyses. The style is applicable to many types of data, but the content and labelling of individual outputs will differ with the data used in its production.

Edited output is the result of user editing of direct output. This is achieved conveniently by the user interacting with a file containing the material which would otherwise have been directly printed or plotted. Edited output may be suitable for publication or for more exacting requirements.

Special outputs are as numerous as there are demands for special information display. Examples include report generation, high-accuracy maps and diagrams and special files to permit user interaction with output information which may be reinput to the machine. The generation of communication tapes for phototypsetting or computer output on microfilm also falls into this category.

THE USER INTERFACE—DESIRABLE FEATURES

It is desirable that the user interface for both input and output approaches as closely as possible the ideal of perfect communication between the user and the geological computer. To aid this primary objective, the interface should provide a high-power ratio so that one unit of user effort generates many units of computer effort, the overall effect is a high-information response gained for the expenditure of little user effort.

The third desirable feature of the interface is that it should accommodate a human feedback loop. The output may instigate fresh ideas in the user's mind which stimulate subsequent (or consequent) input. In this manner the user and the geological computer start to work together to solve problems and to generate ideas.

THE ANALOG OF THE HUMAN MIND

In order that the geological computer shall form a transparent interface between the user and his data, its processing capabilities must be in step with his own. Most logical scientists when confronted by a task break it down into subtasks, each of which is comprehensible and tractable within a convenient time and effort unit. It is essential that this user unit also is taken as the software-processing unit so that one subtask to the user and the unit machine subtask are the same. It then is possible for the user to reduce the overall task to his logical unit tasks and determine analogous unit tasks which the machine can perform for him. The result is that the user controls the logic of the unit tasks, and the machine actually performs the tasks.

Such a philosophy demands that the software-processing units are structured to obtain efficient execution and to minimize duplication of program code. Furthermore, it must be possible for the user to specify that any sequence of these processing units should be executed in order, provided that the equivalent tasks are a logical sequence. It is necessary that the processing units should handle a wide range of data (generalization) and that sequences of processes can be performed (integration).

These requirements of software technology effectively constrain the implementation to involve top-down design, and bottom-up construction of highly modular-structured software.

THE REPRESENTATION OF THE WORLD

In a geological computer the data base represents a small sector of the world as described by geologists. There are several desirable characteristics of the data base if it is to function as some sort of model of the real world and if it is to allow efficient operation of the machine.

It has been noted that a data description provides generality by describing the data to the machine and to the user. Independence of data from the processing software is essential for any synthesis of scientific data because it is impossible to define all possible combinations of data format and structure at the time of writing the software. For similar reasons flexibility in the data base is required so that the same data may be regarded from different logical scientific viewpoints through data-independent structural maps of the data base. The requirement for ad hoc synthesis of different types of data using different structural views precludes the rigid predefinition of the content or structure of the data base.

For the data base to represent the geologist's perception of the world, however inadequately, it must have the capacity to preserve relationships which exist in the real world, and to provide facilities for the user to specify hypothetical relationships which he considers might exist and he wishes to model or test. Currently few data base models seem to provide these facilities; the only one known to the authors is the relational model of Codd (1970).

Thus, the geological computer as defined here would consist of a relational data base, generalized and intergrated software which is broken into logical unit tasks and heavily structured and input/output interfaces which allow good communication with a high-power ratio in favor of the user.

However, no one installation can hold an adequate representation of all recorded geological information and data, and it is doubtful if one installation can provide or support all the software necessary to satisfy the requirements of the user community. There is a well-defined need for system intercommunication involving data, software, and ideas.

SYSTEM INTERCOMMUNICATION

There are four main elements of each system; data descriptions, data files, processing commands, and processing software. It is impracticable to communicate everything between installations and those elements which are relatively immobile are data in large quantities and personnel. The experience of a few geologists is incorporated into the processing software (assuming it models the geological manner of thinking) but it is inevitable that some "live" ideas and expertise must be injected into concepts involving transmission between installations.

The transfer of data files requires a methodology that is independent of data content and structure so that it may form an interface to any generalized data-handling system. Such a methodology (FILEMATCH) has been developed by Peter Sutterlin and the G-EXEC team (Sutterlin, Jeffery, and Gill, 1976).

The transfer of processing commands will require either a standard command language for a network of "local geological computers" or interfaces at each installation to provide the command language of other installations. Both options should be possible, particularly if the process of translation, or the common language, are based on boolean and/or set operators.

The transfer of ideas in this distributed processing context (from the originator to more than one other center at the same time) becomes possible by the use of some form of teleconferencing and the capabilities of such a system (FORUM) have been demonstrated (Vallee and Askevold, 1975). These facilities could extend the geological computer from one installation to a world-wide network. Their implementation, however, depends on the provision of a communications network between installations so that efficient and practical information exchange may be realized.

THE USER VIEW

The purpose of the geological computer is to serve the user, and to help him in his work. The user view of the proposed system should be the primary criterion in its evaluation.

The concept proposed here is that the user communicates with the machine by a set of user-friendly commands which activate an interactive system for aided input and input-editing. The sequence of commands is submitted to the "local geological computer" which attempts to satisfy the requirement. If the total need cannot be satisfied locally the unsatisfied requirement is translated to a common command set or to the command set of a distant geological computer. The command set is passed through the network linking the geological computers.

The command sequence is processed at the distant machine and a derived file of relevant data is returned to the local installation over the network, using a data content and structure-independent file methodology. At the local installation the derived files from remote installations are integrated with the relevant data on the local installation and presented to the user in a form which is a response to the commands input.

Finally the user can interact with this collected data which is relevant to his task using the local geological computer and derive from it the information required.

In the model of a geological computer presented here the user works with his local geological computer installation to solve problems and to perform normal production work. The geological computer system makes available to the user an easy to use and extensive facility for data handling, covering a large world data base and distributed among installations connected over a network.

The user of the geological computer is provided with an extension of his mind and abilities in processing, and a model of geological situations represented by the data within the computer system.

REFERENCES

Codd, E. F., 1970, A relational model of data for large shared data banks: Comm. ACM, v. 13, no. 6, p. 377–387.

Jeffery, K. G., and Gill, E. M., 1976, The design philosophy of the G-EXEC System: Computers & Geosciences, v. 2, no. 3, p. 345–346.

Sutterlin, P. G., Jeffery, K. G., and Gill, E. M., 1976, FILEMATCH; A format for the interchange of computer-based files of structured data: in press.

Vallee, J., and Askevold, G., 1975, Geologic applications of network conferencing; current experiments with the FORUM system, *in* Computer networking and chemistry: Am. Chemical Soc., Sym. Ser. No. 19, p. 53–66.