Computers & Geosciences, Vol. 3, pp. 465–468. Pergamon Press, 1977. Printed in Great Britain

COMMUNICATION OF GEOLOGICAL INFORMATION AMONG DIFFERENT SOFT MACHINES[1]

S. HENLEY
Institute of Geological Sciences, Exhibition Road, London SW7, England

K. G. JEFFERY
Natural Environment Research Council, Central Computing Group (NCCG), Chilton, Didcot, Oxon, England

C. J. FAGE
Institute of Geological Sciences, Exhibition Road, London SW7, England

and

E. M. GILL
Natural Environment Research Council, Central Computing Group (NCCG), Chilton, Didcot, Oxon, England

(*Received* 25 *October* 1976)

Abstract—With the rapid development of geological data-base systems and the advances in computer technology which allow widespread implementation of intelligent 'front-end' software, a consistent approach to the communication of data and information, using machine-independent file structures and formats, is needed. Such an approach will involve files at the data-base level and at the input–output level. Here an 'input–output' level file structure is proposed, to complement the 'data-base' level G–EXEC command file structure.

Key Words: Communication, Graphics, Data files, Data-management system.

INTRODUCTION

The recent development and continuing expansion of geological data bases (Burk, 1973; Robinson, 1972; Jeffery and others, 1976; Cubitt, 1976) has been accompanied by advances in computer technology which allow cost-effective distributed processing. Because of the wide variety of processing requirements, and the similarly wide variety of types of computer, it is logical to map processing facilities into the most appropriate computers. The function of such generalized processing facilities may be defined independently of the available computer hardware. Also requiring definition, however, are the methods of communication between such soft machines, file structures for transfer of data between two programs or systems in the same machine or via a telecomminications link.

Two levels of communication file may be identified: (1) Between two data-base handling systems, commands and data need to be transferred and (2) Between a data-base handling system and a source-file editing system, a wider range of information may need to be communicated including, for example, text reports, and graphic data.

For transfer of data between two implementations of the same data-base handling system, both data and commands need to be communicated. For the G–EXEC system (Jeffery and others, 1976; Cubitt, 1976) the appropriate file is the G-EXEC job, a structured set of commands and data, in which the data component is the G–STAR standard file, a self-describing file whose data content is transparent to the G-EXEC system.

For data-base level communication between different data-base handling systems, no consistent command structure has yet been defined, and there is no agreed storage structure for data files. Thus the structure (interrelationships) of the data must be transmitted in addition to the data values to allow the correct mapping into data bases with different storage structures, and data must be transmitted independently of commands. A file structure which has been developed for this purpose is FILEMATCH (Sutterlin, Jeffery, and Gill, 1977); this was tested at a COGEODATA workshop in 1975 and successfully allowed communication between the SAFRAS (Sutterlin and Cooper, 1972), G-EXEC (Jeffery and Gill, 1973), SIGMI (Kremer, Lenci, and Lesage, 1976) systems, and indirectly with GRASP (Botbol and Bowen, 1975), DASCH (Mundry, 1973), and GEOMAP (Berner and others, 1972).

For communication of information in characters, no equivalent standards have been proposed. The communication protocols which have been defined by computer manufacturers allow the transparent transfer of data, but because of this they do not constitute file structure standards. For graphic data a number of formats have been proposed related to the plotting software packages such as Calcomp and GINO F. For input of commands and data to the data-base handling system, there are well-defined file structures—the data-base level job file structures, wrapped in job control commands appropriate to the host computer.

THE SOFT-MACHINE CONCEPT

A system which insulates the user from computer hardware and low-level systems software provides the

[1]Published by permission of the Director, Institute of Geological Sciences and the Secretary, Natural Environment Research Council.

user with a view of a machine which is not the same as, and is in fact independent of, the computer in which it is implemented. The user sees a soft machine. It is possible to define a number of soft machines of interest to the environmental scientist; one already mentioned is the data-base handling system. Another is the front-end system for source editing, output control, job submission, and teleconferencing, covering operations which are carried out logically on source card-image files. A third soft machine which will assume increasing importance is one which combines some of the characteristics of each to provide an interactive modeling or graphics system. There are implementations of these soft machines within the UK Natural Environment Research Council and the Institute of Geological Sciences as G-EXEC, G-FILE, and G-MODL respectively.

The G-EXEC system (Jeffery and Gill, 1973, 1976; Jeffery and others, 1976; Cubitt, 1976) provides a large-scale generalized, integrated data-base handling facility, with a structured user-interface which is independent of the host machine. G-EXEC is a system which is best suited to a medium-to-large, batch multiprogramming computer, because of the large size of many of the programs, and the large main storage requirements of the data-analysis components of the system.

G-FILE is a smaller system implemented at present on a PDP–11/45 computer, and occupying 45 K bytes (of which 16 K constitute a block of reentrant code common to all users interacting with the system at any time). It performs many of the editing, job preparation, output control, and message handling functions which are carried out appropriately by direct interaction with the user. It is suited thus to computer hardware which is designed for multiterminal interactive, time-sharing operation.

G-MODL shares features of G-EXEC, but provides the possibility of interactive control between (although not within) the execution of logical processes (process programs). Such a system will support modes of operation such as interactive graphics (the generation, for example, of sequences of plot frames representing the rotation of a set of data) or interactive modeling, in which the sequence of events in a simulated geologaical scenario is determined by the user. Such a soft machine requires a computer which is sufficiently fast to give a real-time response to the user for tasks which are comparable in size with G-EXEC process programs, rather than with G-FILE record operations.

A brief comparison of these soft machines is presented in Table 1.

COMMUNICATION BETWEEN SOFT MACHINES

The existence of a number of soft machines, of the same type or different types, requires some method of communication between or among them, if the one to which the user has most convenient access does not contain all the data or facilities he requires.

Table 2 shows the type of communication file required between each possible pair of the three soft machines defined. Little need be said here about the job file structure used for all input to soft machine 1 (data-base system) or for the batch component of input to soft machine 2 (interactive modeling system). For output from either of these soft machines to soft machine 3 (interactive source-data editing) no file structure is defined yet, although there is a definite need to transfer frames or (logical) pages of different types such as graphic data, retrieved files for amendment or editing and reinput to a data-base, pages of text from report generation programs, and from text entries in conference files transferred between soft machines of type 3.

THE G-SEND FILE STRUCTURE

A file structure has been defined for communication from G-EXEC and G-MODL to G-FILE, as well as between two G-FILE systems. This G-SEND communication file allows the transmission in a single packet of any number of text pages, source-data files, and

Table 1. Characteristics of three basic soft machines

Machine	1	2	3
Required response	hr	min	sec
CPU time per transaction	sec or min	sec	Less than 1 sec
Core size per task	100–1000 kbytes	50–500 kbytes	1–100 kbytes
Processing style	Batch multiprog.	Time-sharing multiprog.	Time-sharing reentrant
Soft-machine example	G-EXEC (data-base handling)	G-MODL (modeling and interactive graphics)	G-FILE (intelligent front-end)

Table 2. Communication files between possible pairs of soft machines defined in Table 1

Receiving machine	Sending machine 1	2	3
1.	Data-base level: e.g. G-EXEC job (FILEMATCH if soft machines are different)	Data-base level: e.g. G-EXEC job	G-EXEC job
2.	Data-base level: G-EXEC job	Data-base level: G-EXEC job	G-EXEC job or single record transaction
3.	Input/output level: G-SEND file	G-SEND file or single record transaction	G-SEND file

frames of graphic data, with page–frame types mixed in any order.

The first two records in the file contain header information referring to the whole file. These are followed by the frames or pages of information. Each page is preceded by a pair of header records. The text, file, or plot data follow these and are followed themselves by a trailer record identical to the second record of the header. Table 3 shows the structure and content of the G-SEND file; the information content within frames or pages is independent of the overall file structure.

Table 3. G-SEND file structure

Record No. & columns		Content
1	1–6	G-SEND
	7	Blank
	8–15	8-character job name or conference name
2	1–60	Name-and-address or comments field
	61–68	Project identifier
	69–76	Date in form *dd/mm/yy*
	77–80	User code
Any number of repetitions of the following records:		
3	1–6	Name of program originating this frame or page
	7–46	Name of data file from which frame or page generated
	47–50	Type of logical page (e.g. TEXT, FILE, or PLOT)
	51–52	Advance code for plotter data (0 for over-plot, or 1 for advance plotter)
	53–60	Page–frame number (sequential integers within any one file)
	61–70	Logical *x* limit for plotter (may be omitted for other types of data): floating-point number
	71–80	Logical *y* limit for plotter (may be omitted for other types of data): floating-point number
4	1–80	Any text or blank: trailer record to be used at the end of the page
5 to N–1		Page contents
N	1–80	Trailer record, as record 4
After all pages the last record:		
	1–4	STOP
For a conference file, record 3 has a different interpretation:		
	1–6	Project code of user adding new entry
	7–46	Name-and-address of user adding new entry
	47–50	TEXT
	51–52	Not defined: irrelevant to text pages (blank)
	53–60	Page number within file
	61–80	Not defined: irrelevant to text pages (blank)

COMMUNICATION FORMATS WITHIN PAGES OR PLOT FRAMES

(1) Text is split into logical pages of any number of lines up to 80 characters in length; no lineprinter carriage control characters are included, and vertical spacing is achieved by explicit insertion of blank lines.

(2) Data files are in card image source format as defined for data input to the G-EXEC system (Jeffery and others, 1976; Cubitt, 1976), the G-STAR standard file structure. Data files are transferred one per logical page.

(3) Graphic data are transferred in a high-level communications format which has been developed from a proposed NERC standard plot–file format (Fay and Jeffery, 1975) and a plot–file format D3.1 (Farmer, 1976) used within IGS. The format is based on the Calcomp subroutine PLOT argument list, as a sequence of (x, y, i) sets where x and y are logical plotter coordinates in mm, and i is a code indicating pen position (up or down), symbol or character, or a control function (change pen, end-of-frame). For a symbol or for the first character of a text string, an additional (x, y, i) set is required to define height (x) and angle (y) of the symbol or text; the i code here is set to zero. There is a one-to-one correspondence between plot frames and logical pages in the G-SEND file. The page header contains information of particular relevance to graphic data: the logical plotter limits (xm, ym) and an advance code, set at 0 or 1, which indicates that the new frame is to be plotted in the same area as the previous frame (0) or that the plotter is to be advanced physically before plotting the new frame. It should be noted that these codes allow the mapping of any plot frame or sequence of frames on to any physical plotting device by the use of an appropriate driver program.

(4) Other forms of data may be included in the G-SEND file structure, but have not been defined yet.

TELECONFERENCING AND THE G-SEND FILE STRUCTURE

It is a trivial extension to the G-SEND concepts to use the file structure as a medium to assist teleconferencing. Vallee and Askevold (1975), and its successor PLANET, conferencing when applied in a number of fields, including geoscience. The FORUM system described by Vallee, and Askevoid (1975) and its successor PLANET are versatile in the conferencing modes which they support. Vallee identified five principal styles of conferencing as (1) the notepad with asynchronous entry of notes by unstructured groups of users through relatively long periods (weeks to months); (2) the seminar, addressing a specific topic, again asynchronously through a period of days or weeks; (3) the assembly, an extension of the seminar, but with multiple discussion topics and typically a large number of participants; (4) the encounter, with fewer participants, a shorter time scale (hours rather than days), and synchronous discussion; and (5) the questionnaire with an unlimited number of participants operating synchronously or asynchronously through a structured interface.

It is in the first four styles of conferencing that the G-SEND file structure may be used; the G-FILE interactive system already allows asynchronous conferencing of types (1), (2), and (3). The page-edit function of G-FILE can provide a directory listing of the conference, list any required pages, and allow the user to append new pages. The type (4) mode of conferencing requires shared access to files and thus involve a more sophisticated level of software, but it seems that the G-SEND file structure could be used here with the appropriate software.

CONCLUSION

The G-SEND file structure allows the transfer, in a single packet, of any number of logical pages of data of similar or different types providing a machine-independent medium for communication among separate soft machines.

REFERENCES

Berner, H., Ekström, T., Lilljequist, R., Stephansson, O., and Wikström, A., 1972, GEOMAP—A data system for geological mapping: 24th Intern. Geol. Congress, Sect. 16, p. 3–11.

Botbol, J. M., and Bowen, R. W., 1975, Geological retrieval and storage program (GRASP), *in* Computer based systems for geological field data: Geol. Surv. Canada Paper 74–63, p. 89–93.

Burk, C. F., 1973, Computer-based storage and retrieval of geoscience information: bibliography, 1970–1972: Geol. Surv. Canada Paper 73–14, 38 p.

Cubitt, J. M., 1976, An analysis and management system suitable for sedimentological data, *in* Quantitative techniques for the analysis of sediments: Pergamon Press, Oxford, p. 1–9.

Farmer, D. G., 1976, ESOPP package users manual: unpubl. manuscript, Inst. Geol. Sci., Edinburgh, p.

Fay, C. J., and Jeffery, K. G., 1975, Report of the working group on exchange data formats for graphics information: Natural Environment Research Council Computer Bull. No. 6, p. 12–22.

Jeffery, K. G., and Gill, E. M., 1973, G-EXEC: a generalised FORTRAN system for data handling, *in* Computer based systems for geological field data: Geol. Surv. Canada Paper 74–63, p. 59–61.

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.

Jeffery, K. G., Gill, E. M., Henley, S., and Cubitt, J. M., 1976, G-EXEC system users manual (phase 5): Inst. Geol. Sci. and Atlas Computing Div., Rutherford Lab. Didcot.

Kremer, M., Lenci, M., and Lesage, M. T., 1976, SIGMI: A user-oriented file processing system: Computers & Geosciences, v. 1, no. 3, p. 187–193.

Mundry, E., 1973, Ein dokumentations-und-abfrageprogramm fur schichtenverzeichnisse (DASCH): Geol. Jb., Bd. A7, p. 25–33.

Robinson, J. E., 1972, Economic use of geological data files in exploration: 24th Intern. Geol. Congress, Sect. 16, p. 173–180.

Sutterlin, P. G., and Cooper, M. A., 1972, SAFRAS: a generalised data base management system: unpubl. manuscript, Western Ontario, London, Canada, 71p.

Sutterlin, P. G., Jeffery, K. G., and Gill, E. M., 1977, FILEMATCH: A format for the interchange of computer-based files of structured data: Computers & Geosciences, v. 3, no. 3, this issue.

Vallee, J., 1976, The FORUM project; network conferencing and its future applications: Computer Networks, v. 1, no. 1, p. 39–52.

Vallee, J., and Askevold, G., 1975, Geologic applications of network conferencing: Current experiments with the FORUM system, *in* Computer networking and chemistry: Am. Chem. Soc. Symp. Ser. no. 19, p. 53–68.