M/SET

BUREAU OF MINERAL RESOURCES, GEOLOGY AND GEOPHYSICS

Record 1985/5

RECORD

BMR Regional Seismic Refraction Data - Disc and Magnetic Tape File Structures and Data Formats.

by

B.J. Drummond and D.N. Downie

Record 1985/5

BMR Regional Seismic Refraction Data - Disc and Magnetic Tape File Structures and Data Formats.

by

B.J. Drummond and D.N. Downie

SUMMARY

This record rigorously documents the data formats and file structures used to store digital data from seismic refraction surveys of the structures within the crust and upper mantle. Thus existing seismic data will be documented both for future use, and to facilitate data exchanges between BMR and other Earth-science institutions. The data are initially stored as disc files with 128-word records. Each disc file contains one trace. The first record of each disc file is a header in mixed ASCII, binary, and Binary-Coded-Decimal formats, and contains all of the information about the origin of the trace and the digitising parameters. The second and subsequent records all contain 128 16-bit binary integer samples in two's complement code.

When the data processing is complete, the disc files are archived to magnetic tape for permanent storage. Tape files comprise a 16-bit file identification record and one or more 4096-word data records. The data records are constructed by concatenating 32 disc file records. The first 128 words of the first tape data record contain the header record of the original disc file. The last data record of each file on tape is only long enough to store the remainder of the data, and will therefore be between 128 and 4096 words long. The first record of the first file on tape will be a tape header record of up to 72 ASCII characters describing the contents of the tape. On most occasions, all of the data to be archived will fit onto one tape, but the archiving program is capable of making a multiple- tape archive.

CONTENTS

1. INTRODUCTION

One of the duties of the Explosion Seismology Group within the Division of Geophysics of the Bureau of Mineral Resources is to undertake seismic refraction surveys of the crust and upper mantle of continental Australia. For this purpose, low-powered tape-recording seismographs were designed and built during the late 1960's and early 1970's, with further development continuing until 1983. The recorders were described by Finlayson & Collins (1980). Briefly, they consist of four-channel tape recorders which record frequency modulated seismic signals at two gain levels on two of the channels, a radio time signal from the Telecom VNG standard frequency and time signal transmission on the third channel, and a time signal from a low-power digital clock on the fourth channel.

Processing of the seismic data is undertaken on a playback centre described by Liu & Seers (1982). The playback centre is interfaced to a Hewlett Packard 1000 Series computer so that one or the other of the seismic data channels can be transcribed into digital format. Thus the data can be plotted as record sections, and the digital format also allows further processing using all of the advantages of digital theory, such as trace stacking and the construction of phase-free filters.

During the processing stage, the digital data are stored as disc files. They are archived onto magnetic tape for permanent storage when the initial processing and data display are complete. Reference to the data on tape is by the same file names used to store the data as disc files.

The purpose of this Record is to document the structure of the disc and archive tape files, and the data formats within the records of the files. These must be documented for several reasons:

(i) A large data set is already in existence, and it is steadily growing. However, future technological developments will probably lead to the recording of digital data in the field, with a subsequent change in data format. Thus, the rigorous documentation of the present data formats is necessary for future reference to the present data set.

(ii) With the worldwide increase in interest in regional seismic refraction surveys of large-scale crustal structures, the likelihood of data exchanges between institutions is probable. Data exchanges can be successful only if the data formats are rigorously documented.

2. USEFUL PARAMETERS AND TERMS

(i) One word is 16 bits
(ii) Binary integers are stored in one word in two's complement code.
(iii) The disc files are Type 1 files on Hewlett Packard computers. This makes handling on HP computers simple, although it need not complicate data transfers to other computers.
(iv) Magnetic tapes are 9 track, 1600bpi, phase encoded.

3. DISC FILE STRUCTURE

After analog to digital conversion, each trace is stored in a disc file. Disc files are composed of 128-word records. The first record is a header containing information about the trace, such as shot number and recording station number, and the digitising parameters. It contains a mixture of ASCII, Binary-Coded-Decimal (BCD) and binary integer coded information. The format of the header record is given in Table 1.

The second and subsequent records each contain 128 equally spaced samples stored as binary integers. The time base of the trace can be reconstructed by calculating the sampling rate from the header record (see Table 1, words 66-69, 102 and 111) and multiplying by the sample number minus one. The sampling rate is controlled by a crystal oscillator with a frequency accurate to 1 part in 10^5. Thus, a 60 second trace should be reconstructed with a maximum error of 0.6 ms.

Care should be taken when using the first few seconds of the trace. During the digitising run, data are input to the computer via Direct Memory Access to 512-word buffers. However, because of the computer architecture and the controlling software, digitisation does not always begin exactly on schedule, and the first few samples may be missed. This results in the samples in the first input buffer being offset towards the start of the buffer by several samples. However, in subsequent input buffers, the data are stored correctly. The start of digitisation is always set well before any events of interest so that the first 512 data samples can be discarded if necessary.

The number of data samples is stored in word 112 of the file header record. The number of samples will always be an exact multiple of 128, so that the number of data records will be the number of samples divided by 128. Add one for the header record to calculate the number of records in the file.

4. MAGNETIC TAPE FILE STRUCTURE.

Usually, the data to be archived will fit on a single magnetic tape. The tape format for a single-tape dump is illustrated for the case of two files in Figure 2a. The first record on any tape is a tape header containing up to 72 characters describing the contents of the tape. The second record is a 16-word file identification record. The format of the file identification record is given in Section 4.1. The third and subsequent records are data records. Except for the last record of each file, they are always 4096 words long, and consist of 32 concatenated disc file records. The first 128 words of the first data record contain the disc file header record. The last record of each file is only long enough to store the remaining file records; it will be between 128 and 4096 words long.

Each file on tape is followed by an end-of-file mark, and the end of logical information on the tape (ie. the last data file) is followed by two end-of-file marks. Note that only the first record contains the 72-character tape header record.

For dumps of large amounts of data, the archiving program can make a multiple-tape archive. This is illustrated for the case of a two-tape archive in Figure 2b. The first tape has the same format as the example for a single tape archive (Fig. 2a) except that, at the end of the tape where insufficient space is left for the last records of File X, the last record will contain the characters "END OF REEL 01". This will be written after the reflective end-of-tape mark. Tape 2 will begin with the same tape header record as Tape 1. The second record will contain the characters "REEL #02", and the third and subsequent records will be the final data records of File X. The end of logical information will be two end-of-file marks.

4.1 Format of the magnetic tape file identification record

The file identification record consists of 16 words.

WORD 1 File name*, characters 1 and 2
WORD 2 File name*, characters 3 and 4
WORD 3 File name*, characters 5 and 6
WORD 4 File type; always type 1
WORD 5 Not relevant in this application
WORD 6 Not relevant in this application
WORD 7 File size. Positive, sectors; negative, chunks.
(2 Sectors = 1 Block = 128 Words;
1 Chunk = 128 Blocks)
WORD 8 Not relevant in this application
WORD 9 Security code of the original disc file
WORD 10 Not relevant in this application
WORD 11 Not relevant in this application
WORD 12 Not relevant in this application
WORD 13 The logical unit number of the disc on which the file was created. Important only for restoration onto the creating computer system.
WORD 14 The cartridge number of the disc on which the original disc file was created. Important only for restoration onto the creating computer system.
WORD 15 Creation date of the original disc file
WORD 16 Last access date of the original disc file

* The file name (6 characters) in words 1, 2 and 3 of the file identification record is the same as the file name of the disc file just prior to archival. It may not necessarily be the same name under which the disc file was originally created. File names can be changed easily, and often are in order to avoid the same name being given to two files. The original name under which the file was created is stored in the first 3 words of the first (4096-word) data record.

5. REFERENCES

FINLAYSON, D.M & COLLINS, C.D.N., 1980 - A brief description of BMR portable seismic tape recording systems. Australian Society of Exploration Geophysicists, Bulletin, 11, 75-77.

LIU, Y.S.B & SEERS, K.J., 1982 - A playback system for portable seismic recorders. Australian Society of Exploration Geophysicists, Bulletin, 13, 77-81.

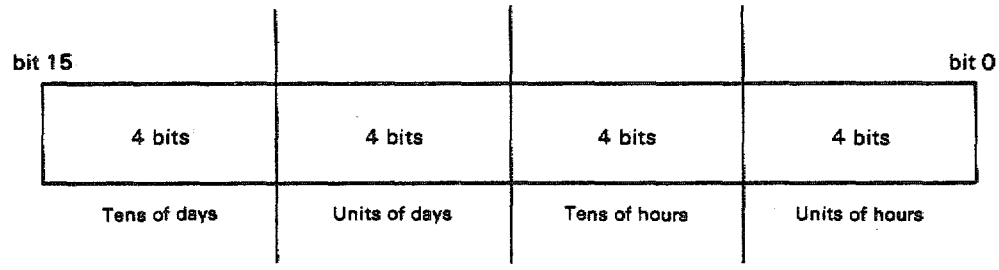
TABLE 1 - Format of disc file header records.

Word Number	Number of words	Contents	Character type	Comments
1- 3	3	Filename at file creation	ASCII	Up to 6 Characters. The filename can subsequently be altered but this name remains unchanged.
4-39	36	Survey description	ASCII	Up to 72 characters which briefly describe the survey on which this trace was recorded.
40-42	3	Survey number	ASCII	6 digits which uniquely define the survey. Usually the date of the first blast recorded for the survey eg. 10 October 1983 becomes 101083.
43-44	2	Shot number	ASCII	Up to 4 digits.
45-50	6	Shot time	ASCII	12 digits, in the format dd hh mm ss.sss
51-52	2	Station number	ASCII	Up to 4 digits.
53-55	3	Distance		Up to 6 digits, specifying the distance from shot to recorder. Should contain a decimal point to specify the the fractional part.
56-58	3	Azimuth	ASCII	Up to 6 digits, specifying the azimuth in degrees from the shot towards the recording station. Should contain a decimal point to specify the fractional part.
59-60	2	Amplifier gain	ASCII	Amplifier gain in decibels, in multiples of 6. Should be an integer (no decimal point), probably left justified. This is the gain of the high gain channel.
61	1	Channel digitised	ASCII	One digit specifying the channel stored in this file: 1 - low gain (high gain minus 24db) 2 - high gain 3 - high gain minus low gain 4 - this is a special run, and the details are given in the message in words 66-101
62-63	2	High cut filter frequency	ASCII	Up to 4 characters, usually including a decimal point. This is the high cut of the filter in the field recorder, not of any filter used in the playback centre.
64-65	2	Low cut filter frequency	ASCII	Up to 4 characters, usually including a decimal point. This is the low cut of the filter used in the field recorder, not of any filter used in the playback centre.

Word Number	Number of words	Contents	Character type	Comments
66-101	36	Message	ASCII	Up to 72 characters which can be used to define unusual characteristics of the digitising run for this trace: (i) If the channel digitised (word 61) is set to 4, this message records the exceptional parameters of the digitising run. For example, if an analog filter apart from the anti-alias filter is used before digitising, which is unusual, the message might read "ANALOG FILTER - 5 TO 10HZ, 48DB/OCTAVE". (ii) If the first 2 characters (word 66) are 'CF', the following 6 characters define a number (F6.4) which, when multiplied by the sample interval, gives a more accurate sample interval. This facility is used occasionally when a field recorder tape drive operates at an incorrect speed. (iii) If the 9th and 10th characters (word 70) contains 'IN', the trace is inverted.
102	1	Playback speed	ASCII	1 or 2 characters, left justified. The tapes are replayed at 4, 8, 16 or 32 times their record speed. This factor must be multiplied by the sampling interval (word 111) to get the interval at which the recorded trace is sampled (see also note (ii) relating to word 66)
103-105	3	Shot size	ASCII	Up to 6 characters, possibly containing a decimal point. Shot size is usually specified in tonnes.
106-107	2	Start time of digital trace	BCD	The time in days, hours, minutes and seconds of the start of the digital data. The format is given in Figure 1.
108-109	2	Stop time of digital trace	BCD	The time in days, hours, minutes and seconds of the end of the digital trace. This is unreliable. The format is given in Figure 1.
110	1	Fractional part of seconds of start time of digital trace	Binary	Words 106-107 contain the start of the digital trace truncated to the nearest second. Word 110 contains the fractional part of the seconds, stored in 100ths of a second; eg, if the fractional part is .45, word 110 will contain 45.

Word Number	Number of words	Contents	Character type	Comments
111	1	Sample interval in milliseconds	Binary	The A/D converter samples the data at this interval. The interval at which the recorded data are sampled is derived by multiplying this factor by the playback speed, and, if necessary, the factor in words 66-69.
112-113	2	Number of samples	Binary	Word 112 contains a 16-bit binary integer number of samples in the digital trace. The number will be a multiple of 512. Word 113 is provided for possible future extension to a 32-bit double-word integer to allow the number of samples to be greater than 32768.
114	1	Security code	Binary	The security code under which the disc file was protected at creation; stored for reference purposes only.
115	1	Cartridge number	Binary	Disc cartridge on which the file was initially stored; for reference purposes only.
116-128	13	Not used at this stage		

Word 106 of file header record



Word 107 of file header record

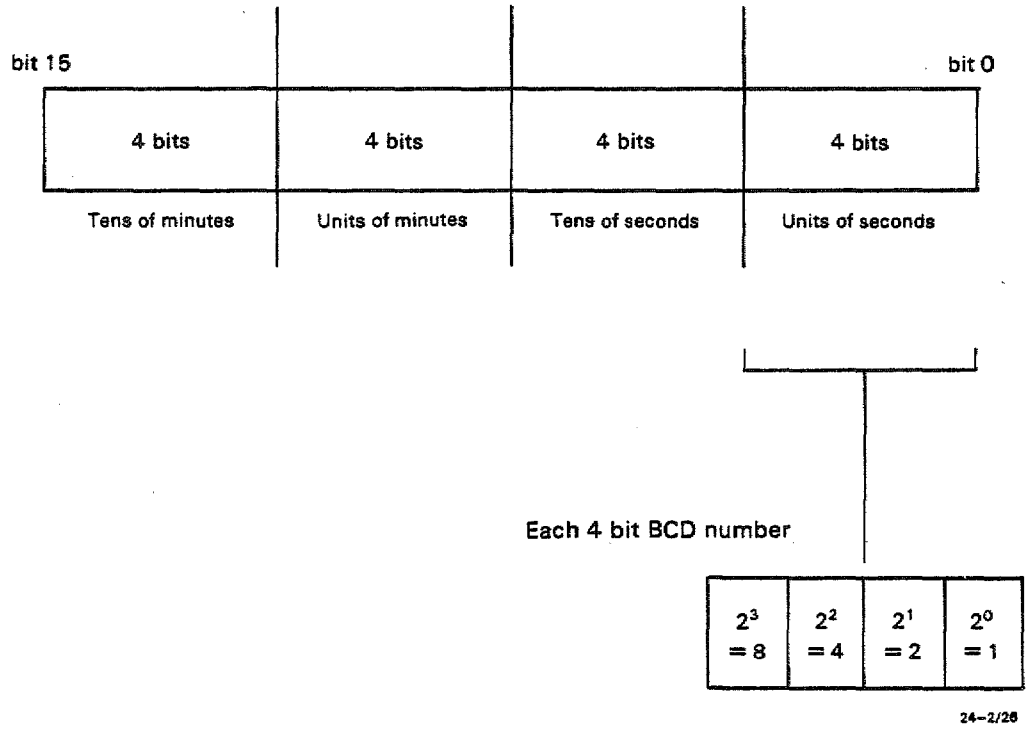
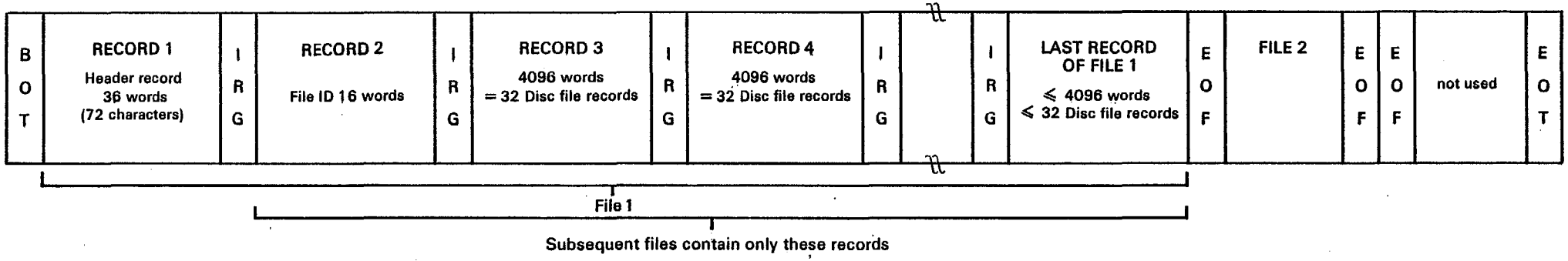


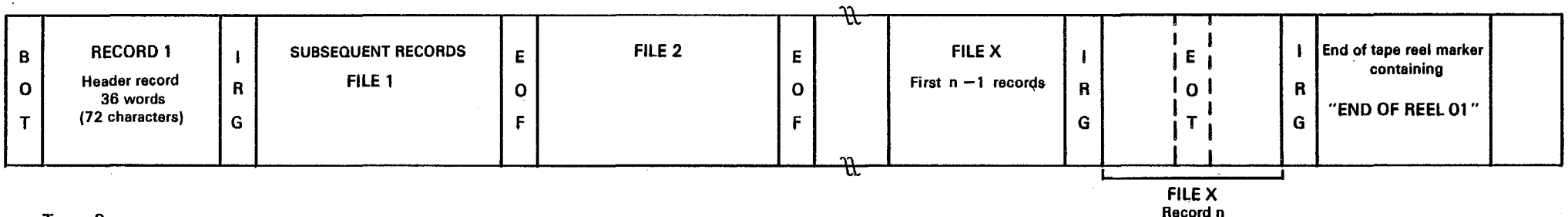
Figure 1. Format of digitising start and stop times.

(a) Single tape archive — 2 files

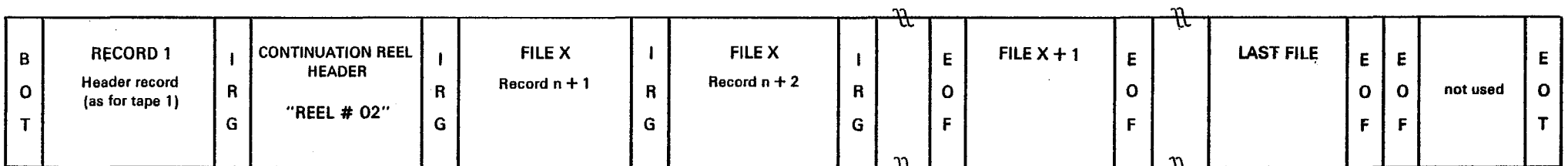


(b) Multiple tape archive — 2 tapes

Tape 1



Tape 2



24-2/27

Nomenclature:
- BOT — Beginning of tape
- EOT — End of tape
- EOF — End of file
- IRG — Inter-record gap

Figure 2. Magnetic tape format for archived disc files.