

SeisCap Duo. Seismic Record Encapsulation.

Eric Keyser

EnCana, Calgary, Canada

Background

Ever since the 1960's seismic data has been stored on magnetic tape. During the 1990's there was recognition that tape would not last forever and data needed to be preserved in a manner that was independent of physical media. In some cases the original physical media devices that the data was stored on were becoming obsolete and it became difficult and expensive to retrieve the data. Companies then began storing data on modern random access media as electronic files (hard disk or optical disks). This move to media independence led to major projects that resulted in millions of tapes to be converted to electronic file formats. Only now are we finding as we retrieve the data that a majority of the work was done without taking Archive Principles into account. Tape has the characteristic of being a perfect self defining record-oriented media. Tape has always had the advantage that it preserved the order of what happened in the field. Field problems are more easily resolved when you know the order of events.

Vector Archives in Calgary was one of the first companies to identify that it was important to preserve order and to keep the data in an original bit for bit form. It is amazing that archive companies in the industry are still converting original SEGD field data to SEG Y for archive purposes. Data from the trace headers is typically lost when data are reformatted and data integrity is impossible to verify unless you repeat the process. Vector's storage model consisted of three files: a binary data stream of all records concatenated together, an index file that defines the start and length of a record, a check sum value and a log file.

Kelman Archives later adopted this index technique and created the Kelman Quartet (KQ) to encapsulate the following seismic formats: SEGA, SEGB, SEGC, SEGD and SEG Y. In addition to both the data stream and index, Kelman provided a log file and their own internal database sum file. The log file became very important in order to identify exactly what happened in the archiving process. A simple byte checksum is part of the index file and used to identify possible file corruption.

Seitel Solutions adopted an almost identical methodology consisting of just three files: the binary data stream, the index and the log file. Seitel simplified the manner how large files were handled with the implementation of a 64 bit file pointer. Kelman had to maintain back words compatibility and implemented a complicated pointer extension as a way to extend their format above the 32 bit file pointer boundaries.

Two years ago I was asked if I could decode a RODE encapsulated tape that came from the Middle East. I spent a day

and quickly came to the conclusion that this was a non-trivial problem. There are valid reasons why commercial programs to decode RODE are very expensive. I came to the conclusion that geophysicists needed to have a better way to archive the vast quantities of seismic data.

Seismic archiving principles

- The original data structure is in no way altered. The binary data records remain in their original record order and all EOF (End of File) markers are preserved in their original location in regards to the data stream. (Multiple data records stored on tape.)
- The original data content is not altered in any way. Information is not added or deleted to the data that is part of the binary data stream. All trace headers and test records remain intact as well as normal field data errors such as parity errors and short records. (Reformatting datasets makes it impossible to adhere to these principles.)
- The established encapsulation format is independent of input data format type; one encapsulation format can support any data format type. (Sega, Segb, Segc, Segy, Segd etc.)
- The encapsulation format supports the input from any data source whether it is 9trk tape, 3480, 3590 media or electronic files. The encapsulation imposes no restrictions that are characteristic to physical media devices. It also has no restrictions in regards to record sizes for record length or number of files exists. Formats that are media based exhibit these characteristics.
- The encapsulation format insures data integrity is maintained for the life of the data. The encapsulation format must implement the use of a data signature in the form of a Check sum Value. Embedded Check sums ensure the data remain unaltered and in their original form.
- The encapsulation format contains the necessary information required to perform data verification. A living verification of the data should be part of the process of using the data to insure integrity.
- The encapsulation format has the ability to track media errors that occurred in the archiving process. Tracking media errors allows users of the data to understand why poor quality data records exist, and ensures problematic data errors are not masked in the archive.
- One single Encapsulation format. As future additional industry data formats evolve, the encapsulation format does not need to be extended or modified to support data structure changes. (Example - Vanguard is not a single format, rather a collection of many formats that depend on the input data type.)

Continued on Page 43

SeisCap Duo

Continued from Page 42

- The archive format supports the ability to log and track the physical archive process used to create the archive file set. Operator's actions that took place during the archive session are recorded. This insures the repeatability of the actions that took place during the archive process and it is useful in establishing the reasons behind errors and in resolving data issues.

SeisCap Record Oriented Encapsulation

I reviewed all the seismic archiving techniques I could find; I looked at RODE, Lacey, Vanguard and others. I felt that the industry needed a technique that followed basic archive principles. It had to be simple and could handle all file formats including those that have yet to be defined. I found that Seitel Solutions came closest to what I was after. All that had to be fixed was to extend the format so that all data types could be handled and to improve the check sum algorithm. C&C Systems in Calgary helped with the final steps of the process and they have written software to reformat data both in and out of SeisCap format.

SeisCap files consist of a pair of files, the data stream and the index. To encapsulate all file types, the first file of the SeisCap data stream consists of the complete file name, file association and description field. The file association is defined so the SeisCap extractor knows what program to use to interpret the extracted file. The last file in the archive stream contains all the log information that was generated by the encapsulation process. The length of each file is contained within the log file so the data could still be extracted if the index file did not exist. The index file contains the robust 16bit CRC checksum algorithm.

SeisCap is recursive. It is recommended that all the shots from a physical tape be prefixed with a picture of the physical tape and then encapsulated. These encapsulated field files are then encapsulated together with the basic data (observer's, driller's, chaining notes, survey, etc). All basic data for a complete 2D or 3D seismic line are located in a single file for archive purposes.

SeisCap details

The first 12 bytes consist of the Version Number, the Number of Files and the Number of records. The numbers are defined in Big Endian (SUN) byte order.

Byte Offset	Format	Description
0	4 byte Integer	Version (SeisCap=4, Kelman=2or3, Seitel=1)
4	4 byte Integer	Number of Files
8	4 byte Integer	Number of Records

Each file is described by 36 bytes of the index record. Only two 64bit numbers need to be read to extract a file or record. Byte offset 12 contains the start location and offset 20 the length of the file. The record index also contains the 16bit CRC check sum and a few fields to identify data.

Byte Offset	Format	Description
0	4 byte Integer	File Number
4	4 byte Integer	Original Field File Number (1-9999)
8	4 byte Integer	Record Number
12	8 byte Integer	Start Location (Offset)
20	8 byte Integer	Record Length (0=EOF)
28	1 byte ASCII	Format codes, a=sega, d=segd, y=segy etc
29	2 byte Integer	CRC 16 bit Check Sum
31	1 byte Integer	Status, normally it's set to 1
32	1 byte Integer	Reserved for future use
33	3 byte ASCII	Optional, 3 letter file suffix

SeisCap in practice

The SeisCap Duo encapsulation technique has now been published. Both the SeisCap creator and the SeisCap extractor have been placed into the public domain. They are available with additional details at the web site <http://www.segy.ca>

So far the largest single encapsulated file is now more than 60gig. The limiting factor seems to be the speed of the network. The time to encapsulate is insignificant compared to the time to copy data. The last record can be extracted just as quickly as the first record. We routinely electronically transfer SeisCap Duo files to our seismic processors.

The final step in the Encapsulation Process is the verification step. The fields shot data are demultiplexed and brute stacks are generated. We have now verified the encapsulation and the original field tapes can be destroyed or left to decay over time.

The future of SeisCap

Current practice in the field is to record seismic data on disk in a buffer and some time later create output products. Normally this consists of tapes (LTO seems to be the preferred for 2006) and DVD's for the basic information. It is recommended that the archive step be created as soon after the data acquisition as possible. All shot data should be encapsulated with check sum information at the field recording truck.

References

This paper would not be possible without the frequent communications over the past several years with Bill Leakey of Seitel Solutions. Bill provided the details for the different encapsulation techniques as they evolved in Calgary. Randy Peters of Kelman Archives provided the details for the Kelman Quartet Index file Documentation. Carmine Militano of C&C Systems provided the programming support for SeisCap and the valuable insights from a Seismic Processing Shop perspective. Arthur Vanderhoek of EnCana provided the details of how to manage more 30 terra bytes of online encapsulated data. **R**