Data Format Considerations for Continuously Recorded Seismic Field Data

by Audrey Hughes

Introduction

With new technologies and advances in seismic data acquisition and processing methods, the oil and gas industry is now challenged by how it can efficiently manage the large volumes of additional data obtained by these new approaches. These challenges are experienced particularly with the arrival of continuously recorded seismic data which produce much larger datasets than the more traditional marine towed seismic methods and which push the current format standards for seismic data.

Traditional data management practices are proving inadequate when dealing with these huge datasets and implementation of new Data Management (DM) protocols should become a primary consideration for companies if they are to maintain the value and regulatory compliance of this new data.

The way this data is managed could impact on a company's performance and it is important that rigorous DM practices are put in place. These practices should be improved to maximise the value of the data and improve overall performance towards data quality, access and indexing systems, preservation, security and governance.

With the increase in use of ocean bottom node and blended simultaneous source techniques, more continuously-recorded datasets are being acquired and processed and there is a need for standards to best manage this type of data. In the UK these datasets are proprietary for only four years before they must be released and there is a need now for a clearly defined standard for storage and distribution of this data type – something that has still to be established.

The multi-components, high fold, long offsets and wide azimuths involved in this type of acquisition can give much improved data quality compared to that of marine towed seismic, but since there is no clear standard for these new field data file formats this article is written to promote the understanding of these issues amongst stakeholders and to suggest guidelines for field data deliverables and the management of these types of data.

Archival and Processing Formats

Archival and storage of seismic field data in the industry has, for the last 25 years, revolved around the SEG-D format. This is a flexible format where multiple data types can be recorded along with a wide range of header information that can be read and employed during the processing of the data. Up until now, industry bodies such as the Oil and Gas Authority (OGA) and Norwegian Petroleum Directorate (NPD) have required that SEG-D be the format that is supplied from the acquisition contractors and stored and released by the oil company. This model works for the traditional towed cable seismic method where the data can be simply broken down into finite periods of time after the airgun array is fired (the shot records). With a fixed number of receivers, the individual shot records are relatively small (megabytes to gigabytes) and are easily handled and understood by data managers.

Continuously recorded methods, such as nodal or Ocean Bottom Cable (OBC) surveys, require a change to this approach as the primary order of the dataset is now the receiver on the seabed and these can be recording data for up to 48 days continuously. During this time, data is being acquired from a source vessel sailing across the survey area firing their airgun arrays at predefined locations but often at very random times. Much of the recording time is spent listening to ambient noise or passive seismic events. While it may be advisable to archive a set of raw field data containing all this information, it would be very inefficient to retain all of this data which does not contribute to the seismic survey (maybe 75-80% of the recorded data volume) for the purposes of processing and distribution upon release. To handle this, the live portions of data are identified from the continuously recorded dataset by cross referencing this with the times of the source firing. A period of a few seconds is copied for every receiver at each shot point event and these data recorded to a new, much reduced data volume. The remaining data is discarded from the processed data volume.

In addition, each receiver can contain 3 motion sensitive geophones plus a pressure sensitive hydrophone to aid the recording of the complete seismic wavefield. As the receivers are free to land in any orientation on the seabed, the data need to be rotated so that the receivers can be considered to all be pointing in a common X, Y and Z azimuth. This processing stage is often applied on board the recording vessel along with the general QC dataset and, having been reformatted from the standard recording SEG-D, needs to output in a different, but still standard, post processing format known as SEG-Y. SEG-Y provides a trace header which contains information about that particular piece of data, including source and receiver numbers, X & Y locations of the source and receiver and many other header values. These headers include the same information as would normally be used in identifying and storing the field data tapes in a database so this format would be suitable for storage and archival of the seismic dataset.

The simplified workflow in figure 1 is used to illustrate the on-board processing from the recorded field format SEG-D to the field SEG-Y (which is the product for further processing).

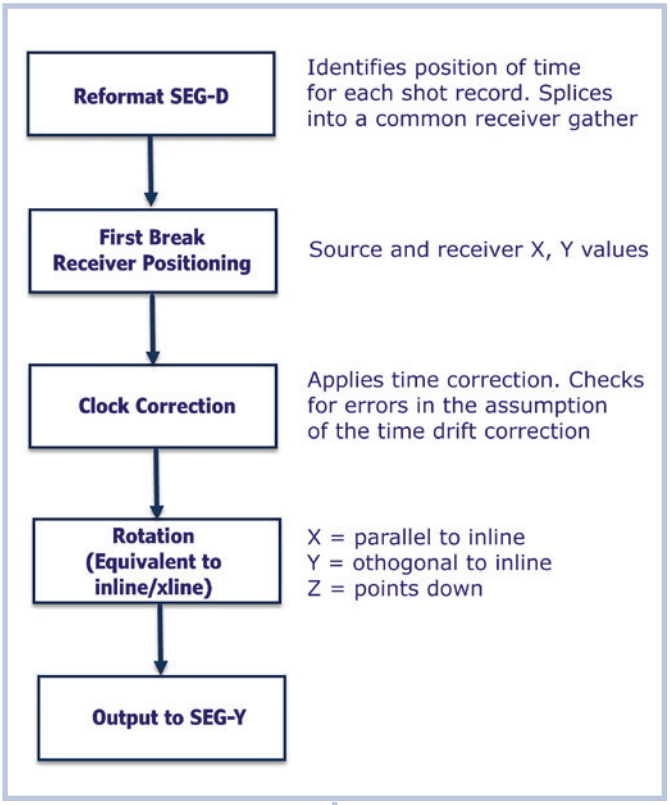

Figure 1. On-board Acquisition Workflow

```
C27 *NONSTANDARD SEGY TRC HEADERS*
C28 ALL DISTANCE AND DEGREE BYTES APPLY SCALAR X0.1
C29 NODEWATVEL:91-92;NODEID:209-210;
C30 FLDSOUXCOORD:121-124;FLDSOUYCOORD:125-128;
C31 FLDRECXCOORD:129-132;FLDRECYCOORD:133-136;FLDRECZCOORD:137-140;
C32 PREPLOTSOUXCOORD:141-144;PREPLOTSOUYCOORD:145-148;
C33 PREPLOTRECXCOORD:149-152;PREPLOTRECYCOORD:153-156;
C34 NODEDEPLOYMENTTIMEMSW:189-192;NODEDEPLOYMENTTIMELSW:93-96;
C35 NOMSOULN:171-172;RECEIVERIDNUMBER:193-196;RECSTNNO:201-202;
C36 RECLINENO:215-216;COMPTYPE:203-204;VESSELPASSNO:207-208;
C37 FLDAZIMDEG:199-200;FLDPITCHDEG:205-206;FLDROLLDEG:213-214;
C38 DEGSCALAR:229-230;DEGUNIT:231-232;
C39 PHASESHFTAPLDVGEO:237-238;TIMESHFTAPLDVGEO:239-240;
C40 END EBCDIC
```

Figure 2. Extract from SEG-Y Card Image Header

The author has performed data analysis on a sample dataset (courtesy of ConocoPhillips (UK) Limited) using the Data Management Utility Suite software tools from Troika International Limited to test the suitability of using field SEG-Y as a release data in comparison to SEG-D. Testing included, but was not limited to:

- Dump of trace headers and extraction of card image header information
- Readability of tapes
- Format checks for compliance and variation from SEG standard
- Mapping of coordinate headers to check geometry consistency and QC with supplied ancillary documents
- Generation of QC displays from SEG-Y data

Reading of the SEG-D tape was made using Troika's Magma transcription software to establish the format and data compliance with SEG standards. With Magma it was possible to analyse the data and establish that the SEG-D had a very long header with 2,365,101 external header blocks which would be problematic for most companies to read. One of the issues with SEG-D is that, despite being described as a standard format, it also has a large amount of flexibility built into it as no two seismic surveys are the same. Each dataset will require some unique feature to be described. This may be a different set of headers or different block size or record lengths. In this case, it is the very large header block that will cause issues.

Analysis of the SEG-Y data was a more straightforward process. The SEG-Y format is a much more rigidly defined standard with each file on tape having a descriptive 'Card Image' header in EBCDIC format of 3200 bytes in length and a standard 400 byte binary header. Each trace in the file then has a standard 240 byte header containing information about the trace to which it is attached. The standards for SEG-Y specify that the first 180 bytes of the trace header contain predefined information at specific locations and that the last 60 bytes are available for freeform header descriptions. In practice, many of the predefined headers have become obsolete or do not apply to the particular acquisition method, so they too are available for redefining.

The Card Image header contains a description of the dataset on tape with acquisition parameters, a description of any processing applied and, more importantly in this case, any deviations away from the standard header layouts. The Card Image header from the first file on the sample SEG-Y tape is shown in Figure 2. This text allows the researcher and anyone processing the data to identify, transcribe and use the key headers further downstream in the flow.

These headers have been defined following application of the survey navigation where receivers and sources have been identified and assigned their X&Y coordinates from the survey information and give the processing company most of the information required to process the data for use in interpretation. Cross referencing this information should be an important QA step when checking data which is to be archived to assure format, compliance and consistency and could provide the basis for an archive database.

The sample SEG-Y dataset was analysed using Troika's Magma and Midi software packages. The data could be checked to ensure that the header values made sense (header names Reclineno & Recstnno should be constant for all traces in a file). Various QC displays were created using Midi to:

- check readability of the supplied data

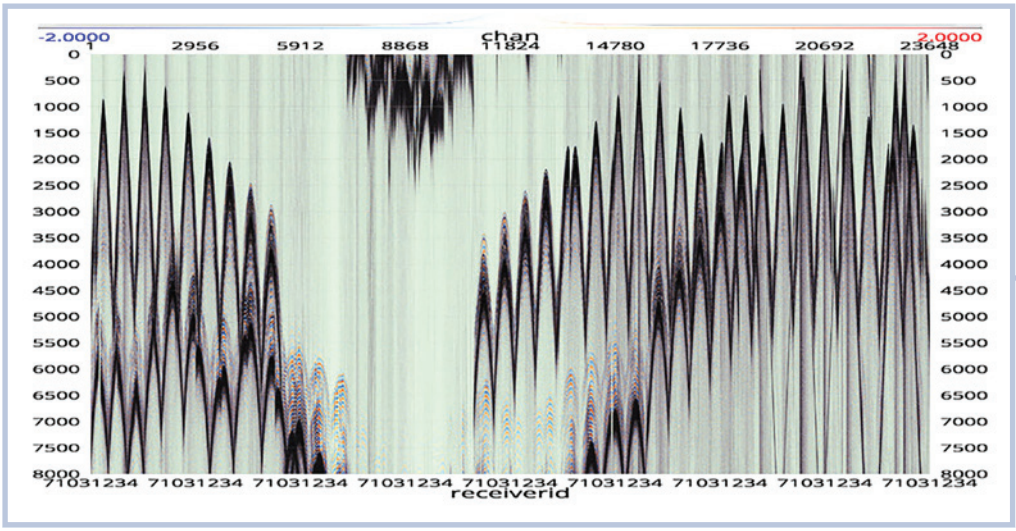

Figure 3. Common Receiver Gather display generated using Midi

- sense check the header information
- QC the seismic data

Figure 3 shows a display of all shots fired into a single receiver (a common receiver gather). This image illustrates the common receiverID header (bottom axis) being a combination of the receiver line number and the receiver station number. This also illustrates the number of sources fired for this swath of data (approximately 23648 – shown on the top axis). The vertical axis is the listening time from the source firing recorded in milliseconds (ms). This information may be considered irrelevant, but it is valuable and the understanding of these displays is an important link between the QC of data and DM. This is a valuable part of the data checking workflow to ensure quality and ease of access should the data be required at a future date.

Summary

Currently SEG-D data is the only form of field data that is recognised by the legislative bodies in the UK, but with the increased use of multicomponent acquisition, it would not be impractical to consider SEG-Y as a release format. The traditional storage format of SEG-D can still apply to the full raw dataset consisting of the complete OBN receiver records and including the periods of ambient noise and passive acquisition when the survey source vessel is not active. As these periods may amount to 70% or more of the total survey time in poor weather, this dataset is inefficient for providing to processing contractors (and for released data) and would be more difficult for onshore contractors to handle. Greater efficiencies can be achieved by editing out the 'non-acquisition' periods and merging the common shot recording periods at the time of acquisition (to provide common receiver gathers as shown in figure 3). Combining this step with receiver rotation would provide a more convenient dataset to provide to contractors and also as a released dataset. As such, this dataset could be considered a post-processed or intermediate seismic dataset and SEG-Y becomes a more viable format for storage and distribution.

Bibliography:

Society of Exploration Geophysicists Technical Standards Committee (SEG) for standards for acquisition and processing formats. http://seg.org/Publications/SEG-Technical-Standards

The author gratefully acknowledges the use of data supplied for the analysis by ConocoPhillips (UK) Ltd.