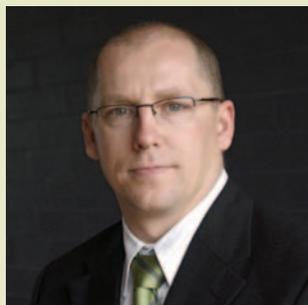




Simple is not always better



Guy Holmes

guy.holmes@spectrumdata.com.au

I was thinking through a project I was working on. It was a project where we had taken some raw seismic field data and processed it in readiness for interpretation. Rather amazingly, the completed dataset was 1/600th of its original size when we were finished. I have experienced this kind of reduction in volume every time I have done this sort of work, but this time it struck me as odd.

Firstly, why should any dataset, after hundreds of hours of work, processing and intellectual input get smaller instead of gaining in size from all of the additional thinking that was put into it? It was kind of like all of my hard work created less instead of more...and, frankly, that was annoying me. It is the job of an interpreter to look for ways to reduce data to information and knowledge (the trees vs the wood etc.) so datasets getting smaller as we move towards an interpretation is to be expected. I guess what was preying on my mind is that in comparison to the original field datasets, I seemed to create less detailed, less informative final data products – shouldn't they be more informative after I put my brain to it?

My gut was telling me that this was less to do with what really should happen with my data and more to do with historical technology foundation stones, laid many years ago, that still drive how we handle our data – even if these foundation stones have no place in today's environment. I decided to dig deeper.

Today field seismic data is generally recorded in SEG-D format. SEG-D stands for Society of Exploration Geophysicists – Format D. The industry started with SEG-A then B, C and now we are on SEG-D Rev 3 8058 32 bit IEEE

demultiplexed (over 20 or so format changes and variations since we started with SEG-A), with the last official update being in 2012. So, when I start a typical project, I usually start with SEG-D. When I am done with my project and the data has been manhandled, processed and all my intellect put into it, I create a format called SEG-Y. SEG-Y, however, is not a smaller more refined version of SEG-D, in fact it bears almost no resemblance whatsoever to SEG-D. SEG-D is so much more complex and data rich that it is usually handled only by specialist companies. In comparison, SEG-Y is simple and handled by almost everyone in the geophysical community (part of the point of the SEG-Y format was to make it simple and easily shared) and with that simplicity you lose a lot of depth, richness and provenance. This is because the SEG-D format is capable of storing more information, more detail and more metadata (more of just about everything really) than SEG-Y.

If you follow the seismic acquisition formats used over time, you will see that they have evolved to allow for the significant advancements in acquisition techniques we have developed. But, the final format of SEG-Y, despite unprecedented advancements in processing technology, is still essentially the same as it was 40 years ago. As we all know, the act of processing data necessitates the need to describe what we did to others to explain the reasoning behind our decision making and the mathematical formulas we applied to the data. But, the final format of SEG-Y has not expanded over time to allow us to comprehensively document our actions so that this information can at least travel with, or be contained within, the data itself. On the face of it, SEG-D appears to have been keeping pace with the evolution of technology given the number of changes that have been made to the format over time, but what of SEG-Y?

The SEG-Y format was created when 3D seismic was not commercially available and techniques like 4D seismic, near field and sea bed acquisition had not yet been imagined. Yet the datasets we are currently creating end up wedged into the same format we created when 24 channel 2D seismic was cutting edge. For me, SEG-D is like a high definition motion picture starring our favourite actors and actresses with heaps of special effects,

CGI and surround sound, and SEG-Y is like a screen capture of a single frame from the end of the movie – and interpreters are expected to use that single frame to determine the plot and final outcome of the movie.

Why would we elect to create our final datasets in a format that has so much less to offer? Well, I think it relates to the following factors:

1. The fundamental design of the SEG-Y format has not kept pace with our technological advancements.

The original field format specifications in the 1970s were designed to take in lots of complex detail *en masse* because in the 1970s computer power and RAM were struggling to handle the volumes of data coming in from the recording instruments. A way to receive data *en masse* and write to recording media as fast as possible, with as much detail as possible, was needed. But, this level of detail and complexity had to be pared back for end users. Datasets had to be created that were of a manageable size and compatible with the computer systems and interpretation technology available. Hence, a massive amount of detail from the acquired data had to be tossed out just to wedge it into the ubiquitous and sharable SEG-Y format. Maybe this was not such a big deal in the 70s when we did not have the advanced acquisition and processing technology we have now. But, the format we use for our final products and interpretations today still provides the same view of the data it did over 40 years ago despite the incredible technological advancements we have made. What if I wanted to 'undo' some of the mathematics applied to the data when it was previously processed and to model the data before the application of that change? Why can't the format itself store what was applied in a more meaningful way, and allow a user to turn it off or to make changes to the mathematical formulas to gain a different perspective of the data? While it is nice to get the data to 1/600th of its original size, it is no longer essential to do so given that our capacity to handle large datasets has expanded dramatically since the 70s. Ironically, it seems one of the major limitations in getting more from today's data is

not a lack of technology, but more that we seemingly remain satisfied with using the same 40 year old format to store our data.

- 2. The heavy adoption of the formats in the industry has created reluctance to change.** The user community adopted the practice of simplification quite rapidly. Once adopted, SEGY became the norm and changing it is not going to be easy. The SEGY format was originally created in 1975 and was officially updated only once – in 2002. The update was not at all significant in terms of differences in the data format – more just a small extension to the original format. How could a format written in 1975 still be considered appropriate for today's complex geophysical datasets? Could SEGY have really been so good that it does not need to be improved? My belief is that in 1975 it was certainly a great

format but, truthfully, the format has not kept pace with our technological advancements. Seismic acquisition is like paying the cost of producing a big budget motion picture on HD DVD, and our acquisition formats happily handle that level of detail, but when interpreting the data we seem quite prepared to watch that movie on a 1975 black and white television set. Why have we not taken bold steps to improve the situation?

- 3. The mounds of technology built to use the data we create would need significant changes and possibly a more 'open source' approach.** Once the use of the simple format became the norm, massive amounts of technology were then built to use the data. Desktop analysis, processing and interpretation systems spread and were rapidly adopted. All of these systems are happy to store your detailed project information including the

filters you applied, the decisions you made and the mathematics you used, but the SEGY format itself does not allow you to retain that detail. Why not? Is it that a massive change to the simple format of SEGY would create havoc in the software development community? Would this change possibly break our dependence on the need to retain particular processing software and hardware platforms?

I have tremendous respect for the Society of Exploration Geophysics and the people who wrote the data formats and the software that handles complex datasets and this in no way is a swipe at their efforts. But, I have to ask, are we keeping pace with technological change or is it time for a rethink? I would appreciate any feedback readers can muster on how other formats used in the geophysical industry have evolved over time.

Introducing a new Global Partnership in Minerals Exploration



The leaders in AEM solutions for hydrogeophysics and the proven specialists in mineral exploration & targeting

ASEG-PESA 2015, BOOTH 110

aarhusgeo.com

newexco.com

