



# geoconvention

Calgary • Canada • May 7-11

# 2018

## Google your way to maximising geoscientific value

*Marc J. Boulet, Data Scientist, P. Geoph., Cenovus Energy Inc.*

### Introduction

Although there is an exponential increase in data collection within oil and gas companies, access to the data for meaningful subsurface interpretation is constrained by the limited toolkits many geoscientists have at their disposal. We propose some methodologies and open-source software tools that can provide more efficient access to structured and unstructured data, thereby enabling geoscientists to make full use of this valuable resource.

### Theory and/or Method

As we enter an age in which every aspect of our oilfield operations is measured, there has been a corresponding explosion in recorded data that is potentially available to the modern geoscientist. However, the multitude of generated data types and formats don't often lend themselves to easy categorisation or access. Well logs, seismic data, XRD data, seismic observation reports, etc. can be stored or organised in various ways, depending on a vendor's or client's ever-changing standards. And while some data can be stored in a relational or object database, Blinston and Blondelle<sup>1</sup> estimate that 80% of geoscientific data is stored in a semistructured or unstructured form.

Typically, there are two methods in which data is stored and accessed. The majority of companies opt for manual curation, in which content is categorized "by hand" according to a certain set of principles. These principles may be rigorous if an organisation employs a comprehensive data management and database storage strategy, or it can be an ad hoc "democratic" process in which individuals place data in directories of their choice, with little oversight. Either approach, however, requires a lot of human intervention, and is fundamentally at the mercy of changes in company organisation and strategy.

The other method is similar to the approach that has made Google an omnipresent force in our daily lives. Access to the early internet was dominated by many companies who provide manually-curated lists of websites<sup>2</sup>. However, this solution was superseded by Google's search and indexing algorithms, which rendered those companies irrelevant and made searches for any topic trivially simple. As an analogue, we propose that an oil and gas company can deploy open-source tools such as **Elasticsearch**, which indexes terabytes of data, regardless of how it is structured or stored. These tools also allow indexing of document contents, including images or slide presentations, if open-source OCR (optical character recognition) solutions are also incorporated.

As a result, a geoscientist can "google" their own datasets and retrieve lightning-fast results, instead of relying on manual or inefficient file-system searches. For example, a geoscientist can identify all play-specific presentations or formation-specific thin-section reports within seconds or minutes, instead of hours or days. If full data categorisation is desired, machine learning algorithms can be employed to automate the process<sup>1</sup>. While some technical knowledge is required for setup, the capability to learn and deploy these open-source tools is easily found on-line and well within the reach of any coding-competent geoscientist.

## Examples

*Note: Cenovus has just begun to index its internal data storage, so these examples are illustrative only, based on initial tests. By the time of Geoconvention, we will have ample quantitative examples of time and workflow efficiency gains due to our indexed system implementation.*

Cenovus began a file indexing initiative in early 2018, after many attempts to manually organise subsurface data, which failed due to lack of resources or suitable data management solutions. In 2016, an attempt to categorize geoscientific data into a structured database has proven successful; however it is still an on-going process that requires a number of full-time staff to administer. The indexing initiative tackled shared network drives containing ~50TB of semi-sorted unstructured data, consisting of Excel documents, PDFs, images, LAS files, and PowerPoint presentations, among many other formats.

By allowing geoscientists access to the indexed data, searches are no longer exercises in missing results or frustration. The index can equally find a five-year old economic model from a long-gone exploitation engineer or a core analysis report embedded in a misnamed file. This indexing system has also proved a real time-saver in enhancing the structured database. Of course, indexes are blind to data-types, so the process can easily incorporate engineering or financial data in addition to geoscientific data. The initiative has proven so successful that it has been extended to other departments, including marketing.

## Conclusions

The exponential increase in data collection is poised to revolutionize the way oil and gas companies operate. However, geoscientists shouldn't have to rely on outdated tools and methodologies to fully unlock the value of this data. Modern indexing software can be employed to empower geoscientists to provide the best possible subsurface interpretations, based on all the available data.

## Acknowledgements

Tamer Salama and Sheldon Wall, Cenovus Data Science Group  
Jessica Galbraith, Sr. Decision Analyst, P. Geoph., Cenovus Energy  
Blair Halter and Steven Milbradt, Cenovus Geoscience Centre of Excellence  
Kirk Duval, P. Eng., Staff Reservoir Engineer, Cenovus Oil Sands Production

## References

1. Kerry Blinston and Henri Blondelle (2017). "Machine learning systems open up access to large volumes of valuable information lying dormant in unstructured documents." *The Leading Edge*, 36(3), 257–261. <https://doi.org/10.1190/tle36030257.1>
2. Season 4 of the AMC television show *Halt and Catch Fire*, available on Netflix, provides a fascinating fictional glimpse into the time of the early Internet and initial attempts to categorise the web.