Submission code: MWBGQ8PXVL

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

Is Our Geophysical Data Ready for AI?

Jess Kozman (Katalyst Data Management)

Is Our Geophysical Data Ready for AI?

Introduction

One trend in recent geophysics workflows is the rush to use new data-driven technologies such as machine learning and generative AI algorithms for real-time monitoring and reservoir management. Use cases have shown these applications have the potential to enhance efficiency, reveal new insights and patterns, and apply high performance compute to massive datasets. The renewed visibilty of these increasingly complex subsurface data volumes has also led to a new understanding of the necessity of properly storing, managing, cataloging and curating geophysical data to make it findable, accessible, interoperable and reusable (FAIR). This is especially true as legacy data is retrieved and re-purposed to optimize a transition to sustainable low-carbon energy projects.

Method and Theory

We deployed and tested optimum industry accepted practices for large volume digital data management on cloud platforms. We used selected public domain metadata and behavior patterns from a database of over 100 petabytes managed under confidentiality and service agreements with multiple global energy and resource operators. The goal was to evaluate our own compliance with industry FAIR data implementation profiles, and to to compare with sources of open-file, public domain data. The selected datasets include geophysical data used for time-lapse and continuous fiber optic real-time passive monitoring, integration with well and production data, monitoring of reservoir injectivity units for fluid dynamics and rock mechanics, and monitoring of induced seismicity from enhanced recovery operations. We also evaluated the deliverabilty and consumption of mandatory metadata used to identify candidate geophysical data sets for compute intensive processing operations such as full waveform inversion. We then tested our ability to train and use machine learning models and generative AI agentic frameworks using the aggregated and abstracted metadata relationships, to reduce data decision latency, or the time spent by geophysical users finding and accessing their data.

Results and Conclusions

We were able to quanitfy and measure the compliance of a sample demonstration database with publically released FAIR implementation profiles and determine how “AI ready” various geophysical data sets are likely to be when obtained from fully and robustly curated datasets, or from public domain portals. More importantly, we were able to identify bottlenecks in the preparation of data for reuse supporting diversified energy and decarbonization projects, and recommend data management mitigation procedures that will enhance the value of those targeted datasets when used as input for AI enabled workflows.