

Unlocking the value from the 50 years' old Exploration Data

by Håkon Snøtun, Project Leader, AGR Software



Håkon Snøtun

Project Lead

hakon.snotun@agr-software.com

The Unstructured Data Challenge was launched last year with the aim of proving that modern data and information sciences could extract half a century old unstructured data which could be used to create context and clarity by combining it with the structured data.

CDA (Common Data Access), the subsidiary of Oil and Gas UK (established to facilitate the sharing of well and seismic data by the oil and gas industry), launched access to their unstructured data in the summer of 2016. CDA wanted to work with a small number of vendors to see how they could unlock the knowledge in CDA's vast data repositories to help the search for hydrocarbons. AGR Software team welcomed the challenge along with eight other contractors.

As part of the CDA challenge, AGR's Software team were given more than 50 years' worth of data, or as a comparison, 3.5 Terabytes of files, logs and images in a plethora of formats and quality.

AGR carried out the project using its own iQx™ data management software to tackle the CDA data, looking specifically at final well reports, many of which were handwritten with no consistent structure.

AGR Software's main focus has always been to make available data accessible; so the de-

velopers started out defining the structured data, finding formation tops and surveys for more than 5,500 wellbores, and parsing well-bore logs to utilise drilling data.

It was found that the data in the Common Data Access was structured in much the same way as on the Norwegian Continental Shelf. Being able to complete the data set for both the British and the Norwegian side is of great importance, since geology is the same despite national borders. Too often we see people use less relevant wells in the same sector rather than the ones across the border because the data is not readily available.

When planning wells, we find that structured and historic data about similar wells is of tremendous benefit in finding trends, making predictions about the area, equipment, time and cost. When anomalies in the data are found, the planning team often spends a lot of time going through verbose final well reports to find if the anomalies arise due to data errors or whether they represent a risk for the project.



Source: AGR



AGR Software developers

That is why we wanted to contextualise the data by presenting the relevant information in the application itself. The team at AGR Software started out looking at the final well reports using OCR to make them machine-readable, then used open-source tools like Lucene to index and make the data searchable. They then began looking for the relevant headers to be able to extract the relevant data such as operational summaries, experiences and risks. Although some of the data was saved as scanned pdf, the team were able to extract value from quite a number of files. When combined with the structured data, it is much easier to understand the context of the data.

CDA did not only want us to create a solution, but gave us an opportunity to define what we wanted to explore, enabling us to think of data

in a new way. We were also fortunate enough to present our ideas and findings not only for CDA, but all the other companies that participated in the challenge. This community had approached the challenge in different and interesting ways, which gave CDA great insight not only into the value of their data, but also novel ways to apply this knowledge.

The results of the work undertaken and findings were presented during a workshop hosted in Aberdeen in late November. A short summary of all presentations delivered at the workshop held after the Challenge can be read here (<http://cdal.com/index.php/2016/12/19/proceedings-now-available-cda-ecim-joint-workshop-on-digital-dividends-from-subsurface-data/>).

