

A New Approach to Manage Unstructured Exploration Data

Mohammad J. Hakami¹, Parwez Sheik¹, Mario Dourado¹, Walid Kaskas¹, Hussein Huwaidi¹

¹Saudi Aramco, Dhahran, SAUDI ARABIA

ABSTRACT

Unstructured data refers to information that neither has a predefined data model nor is organized in a predefined manner; examples of such data are documents, spreadsheets, presentations, images, etc. In Saudi Aramco Exploration Department's repository for unstructured data, more than three million exploration-related documents are stored and hundreds are added on a monthly basis. It is a major task to manage such a large volume of documents while also ensuring the quality of metadata necessary to assist Exploration users in finding the information they need in a timely matter.

A major challenge facing exploration geoscientists is to reduce the time it takes to acquire all the documents related to their projects before they start producing new results. The main reason many documents are not easily searchable with the available search tools is not the tool's search power, but rather the lack of metadata needed to identify and then retrieve those documents. In an attempt to improve the experience of both the producers and consumers of exploration documents, the Exploration Department implemented a new solution to ensure a smoother uploading and querying of upstream unstructured data.

The new interface displays a structured folder format based on the company's Exploration and Production (E&P) taxonomy, which is familiar to Aramco geoscientists. This format will facilitate the capture of documents by allowing them to be dropped directly into the subject-related folder. As a result, when a new document is loaded, it will inherit all the metadata attached to the folder and its parent folders. Additional metadata may be embellished and QC'ed by assigned document controllers. This scheme is a vast improvement in document metadata capture because it overcomes previous issues with missing and incorrect metadata.