Managing Data Quality and Trustworthiness Across Different Organizations

Philip Neri, Ross Philo
Energistics

9.29.2020 - 10.1.2020 – AAPG Annual Convention and Exhibition 2020, Online/Virtual

Abstract

The digital revolution of E&P the past 5-10 years has seen a surge in data volumes feeding into increasingly multi-disciplinary workflows. Datasets are a mix of inputs such as real-time feeds from field activities, dataset subscriptions, active project repositories, archived datasets and 3rd party collaborations. At all steps along a workflow, staff need to be comfortable with the data's reliability before proceeding with processing, analysis or decision-making activities. Typically, this verification and validation gets repeated by each analyst in the workflow, which is clearly incredibly inefficient and represents a significant amount of lost time. A better approach would be if one could transmit the outcome of prior verifications through each stage of the workflow, or between partner organizations, so that this knowledge can be shared. This is not trivial knowing that the definition and criteria to assign a "fit for purpose" status to data varies from one organization to another. Two recent initiatives can help to address this situation. One is a standard format for metadata that captures and transmits the fit or unfit status of a data item together with the criteria used by the organization's data quality policy. A policy consists of a set of rules established by the organization performing the qualification. If all criteria are positively satisfied, the data is deemed fit. Making this information available to a receiving party allows professionals to compare the sending organization's criteria with their own, and either accept the data with its quality information, or identify policy differences significant enough that the receiving organization needs to perform their own screening. The second relevant initiative to reduce data verification workloads consists in providing comprehensive structured metadata along with the actual measurements or results such that if a verification is deemed necessary abundant information is

available to perform a quality assessment efficiently. Absent such metadata, the verification process becomes excessively time-consuming mainly due to the efforts expended searching for information about the data through unstructured documents such as reports or scans.