Submission code: LV5ZLDQRG4

Streamlined Importing of Interpretation Data via Data Science Programming

Marcos Alex Silva de Oliveira (SLB), Daniel De Marins (SLB)

Streamlined Importing of Interpretation Data via Data Science Programming

Abstract

Data management plays a critical role in the efficiency and reliability of geological and geophysical (G&G) interpretation workflows in the oil and gas industry. As subsurface data is generated and interpreted across various platforms, such as those used for well correlation, structural modeling, and stratigraphic analysis, there is a pressing need to ensure seamless integration and standardization of diverse datasets. These platforms often operate with proprietary or incompatible data formats, which creates bottlenecks when transferring, reformatting, and importing interpretation elements. Manual handling of these datasets not only introduces potential for errors but also consumes significant operational time, especially when dealing with high-volume project environments.

To address this challenge, this work proposes a reproducible and scalable data science workflow that utilizes Python programming to automate the transformation, validation, and import of interpretation data across multiple platforms. The methodology is centered on building flexible scripts that adapt to varying input formats such as text-based markers, horizon files, well logs, and polygonal objects, converting them into the target format required by each interpretation environment. This kind of workflow also incorporates quality control routines and metadata enrichment steps to improve traceability and compliance with internal standards. The approach enhances interoperability while maintaining data fidelity and contextual relevance, critical for reducing uncertainty in geological models.

The use case presented demonstrates how this data science-driven workflow can be deployed to reduce processing time, standardize naming conventions, and improve overall data consistency across subsurface projects. By integrating Python-based automation with multi-platform data pipelines, the workflow achieved a reduction of over 90% in manual preprocessing time, while increasing data reusability in both regional and prospect-level interpretations. The adoption of this method also enables faster onboarding of new data deliveries and supports version control practices, essential for collaborative interpretation environments. These results highlight the value of applying programming and automation strategies to enhance operational efficiency in digital geoscience contexts.

This work contributes to the growing field of computational geosciences by demonstrating how modern programming tools and open-source libraries can be integrated into the traditional G&G workflow landscape. It supports broader digital transformation initiatives in the energy sector by offering a robust, low-code-compatible solution that bridges the gap between raw data and end-user analysis. Future applications may explore the integration of cloud-based repositories and real-time data validation APIs to further streamline the process. This reinforces the relevance of interdisciplinary skills in geoscience teams and the importance of scalable solutions for handling increasingly complex subsurface datasets.

Introduction

The increasing complexity of exploration and production (E&P) projects in the oil and gas industry demands more efficient approaches to geoscientific data management. In interpretation environments, such as those used for stratigraphic correlation, structural modeling, and reservoir characterization, the integration of data from various sources and formats remains a critical challenge. Standardizing and organizing this information are essential to ensure consistency in geological models and reliability in subsurface interpretations, particularly when working across multidisciplinary teams and multiple software platforms.

In recent years, the oil and gas industry has increasingly adopted emerging digital technologies to tackle ongoing challenges related to data integrity, operational efficiency, and decision-making under uncertainty. Numerous studies have highlighted the critical role of robust data quality management frameworks (Kang et al., 2023). Additionally, Ahmad et al. (2022) presented real-world case studies demonstrating the application of blockchain-based systems across various oil and gas scenarios, while also discussing key technical and social challenges that limit broader adoption.

Technical literature underscores the value of automation and digital workflows to address these challenges. VELASQUEZ et al. (2019) demonstrated that systematic workflows for geological data transformation and integration can significantly reduce manual effort while improving the reproducibility of results. Similarly, digital methodologies in geoscience not only enhance operational efficiency but also democratize data access and accelerate technical decision-making.

LOPES et al. (2023) noted that one of the major obstacles in interpretation data workflows is the fragmentation of sources and the lack of a unified data structure, which can hinder traceability and reduce interoperability across platforms. LOPES et al. (2023), in their investigation of geological data standardization strategies, reinforced the need for adaptable solutions capable of interpreting multiple formats while ensuring compliance with internal metadata and naming conventions.

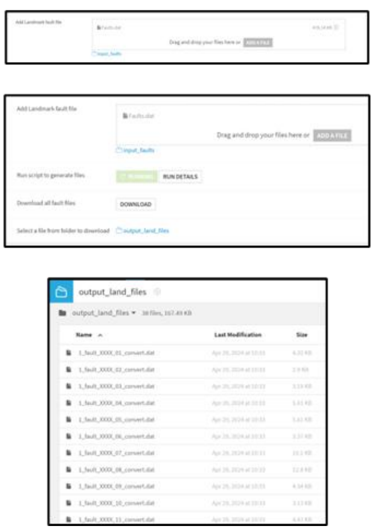

Figure 1: Project created with the step-by-step access to the workflows.

In this context, the present work introduces a framework aimed, presented in Figure 1, at optimizing interpretation of data import workflows by reducing data loading times, eliminating repetitive manual steps, and enhancing the overall quality of integrated datasets. By implementing modern Python techniques in a low-code environment for data transformation, validation, and quality control, this study contributes to the ongoing digital transformation in the energy sector and supports more agile, consistent, and scalable interpretation workflows.

Data Science approach for Exploration Data in multi platforms

The management of subsurface exploration data across multiple interpretation environments poses significant challenges due to diverse data models, file formats, and software-specific requirements. A data science–driven strategy provides a unified framework for handling these complexities by leveraging algorithmic techniques to ingest, transform, and validate heterogeneous datasets. As noted by WANG et al. (2023), establishing systematic ingestion pipelines ensures that raw exploration inputs, such as well logs, horizon picks, and fault traces, are automatically converted into standardized schemas compatible with each target platform. By applying data wrangling libraries and rule-based mapping tables, exploration teams can reduce manual formatting errors and guarantee consistent geospatial positioning, irrespective of the interpretation software used (BRULE, M., 2010).

Once raw inputs are ingested, data science methods facilitate the enrichment and quality control of exploration data. For example, outlier detection algorithms and statistical consistency checks can be applied to well-log measurements to identify anomalous readings before importing into modeling environments (WANG et al., 2023). Similarly, data provenance mechanisms, such as automatic metadata assignment and version tracking, reinforce traceability when multiple geoscientists work on the same regional model.

Interoperability across platforms is further achieved by adopting modular transformation routines that encapsulate each platform's import/export rules as discrete functions. Dransch, D. *et al.* (2022) highlighted that such modularization allows teams to update or extend conversion logic (e.g., adding a new attribute field) without disrupting the entire workflow. Also, Dransch, D. *et al.* (2022). corroborated this by showing that modular pipelines accelerate onboarding of new data types, such as polygon-based geological interpretations or time-indexed seismic horizons, thereby increasing operational agility.

Recent studies have also explored the integration of data science with cloud-based storage and API-driven frameworks to enable real-time synchronization of interpreted results, further reducing loading times and maintaining a single source of truth (BRULE, M., 2010). In parallel, other research has focused on digitalization strategies within the oil and gas industry to enhance data acquisition and analysis processes. These efforts leverage computing advancements such as cloud infrastructure, cognitive computing, and real-time algorithms to support advanced analytical and numerical techniques (Al-Rbeawi, 2023).

Method

The workflow, as shown in Figure 2, was designed to operate through a dedicated application interface, where users can upload input files using the "Add a file" option. Once the data is submitted, the application executes a backend script within Python environment deployed on the cloud platform, triggering automated transformation and processing steps. After execution, the generated output files are prepared for export. For data transformation tasks, users also have the option to download processed files directly to their local machines, enabling further analysis or manual validation when necessary.

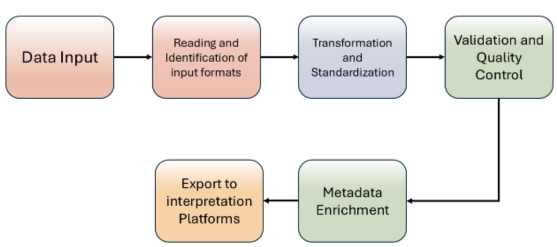

Figure 2: Methodological workflow used in this project, illustrating the process from data input into the platform to the final delivery of interpretation environments.

Results

The automated workflow significantly enhanced the process of importing interpretation and well data. Manual tasks previously required for data loading were reduced through automation, improving both the consistency and speed of operations. In addition, the solution allowed for the development of reusable data flow structures within customized applications, increasing flexibility for different project scenarios. This modularity supported a more streamlined collaboration between data management and technical teams, aligning processes and improving operational continuity.

One of the key outcomes of the project was the successful automation of several geoscientific data types. Interpretation elements such as faults, seismic horizons, and structural surfaces were processed automatically and formatted for seamless ingestion into interpretation platforms. These datasets often require complex spatial alignment and metadata consistency, which were handled through predefined rules and validation steps integrated into the automated workflow.

The well data component included directional surveys, wellhead locations, and LAS log files. In addition, well markers and lithological descriptions were processed and imported using structured pipelines, which reduced errors and improved data quality across interpretation environments. The ability to perform data analysis during the transformation process, enabled by the data science environment, ensured better traceability and performance monitoring. These enhancements contributed to faster project startup times, increased reliability of subsurface models, and broader scalability for future applications, allowing the time the following results:

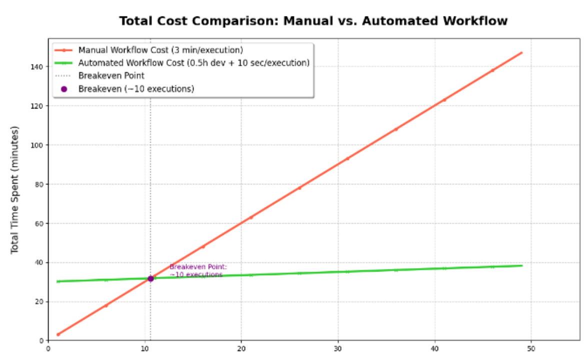

Figure 3: Comparison of total time cost between manual and automated workflows for interpretation data import.

1. Time saving related of **90% of interpretation data and well data** importing process in Petrel Projects, as demonstrated in figure 3.

2. Possibility of executing new tools, data flows and building flows in applications.
3. Total of **21 workflows** into the Python environment project.

Conclusions

This work presents a transformative, reproducible, and scalable data science workflow leveraging Python to manage geological and geophysical interpretation data. By automating critical processes like data transformation, validation, and import, our methodology directly tackles the inherent inefficiencies and potential for human error found in manual data handling across diverse interpretation platforms.

The implementation of this automated workflow delivers tangible benefits. It boosts execution time efficiency, with our presented use case demonstrating a reduction of over **90%** in manual preprocessing time. This automation also significantly minimizes human error, leading to demonstrably higher data quality and consistency. Furthermore, the solution enhances data availability and interoperability across varied platforms, facilitating seamless integration and rapid deployment of vital data to systems used for correlation, structural modeling, and stratigraphic analysis.

Acknowledgments

This scientific publication was developed with the support of software provided by SLB, which has been operating in Brazil since 1949. SLB is a global technology company committed to advancing technological innovations for industrial decarbonization. The donation of these resources aimed to promote scientific production in support of sustainability.

References

Dransch, D. *et al.* (2022). Data Analysis and Exploration with Scientific Workflows. In: Bouwer, L.M., Dransch, D., Ruhnke, R., Rechid, D., Frickenhaus, S., Greinert, J. (eds) Integrating Data Science and Earth Science. SpringerBriefs in Earth System Sciences. Springer, Cham. https://doi.org/10.1007/978-3-030-99546-1_5.

BRULE, M., Fusion of Engineering and IT Transforms the E&P Business, Journal of Petroleum Technology, Vol. 62, Issue 04 (April). SPE-0410-0032-JPT, 2010.

VELASQUEZ, Cesar; OLEA, Ivan; CHEUNG, Charles; ROUNDTREE, Russell. Data readiness and standardization for data analytics studies. In: INTERNATIONAL CONGRESS OF THE BRAZILIAN GEOPHYSICAL SOCIETY – SBGf, 16., 2019, Rio de Janeiro.

WANG, Xuejun; MAO, Jun; JIANG, Min; WANG, Yuyang; SHU, Xing; TONG, Yuanyuan. Integrated management technology of geophysics and geology data. *Advances in Engineering Technology Research – ICISCTA 2023*, v. 7, 2023. ISSN 2790-1688.

LOPES, Christiano; AQUILINO, Igor; ROCHA, Rafael; PEREIRA, Sebastião; FERNANDES, David; RIBEIRO, Leandro; MIRANDA, Frederico; MICHELON, Diogo; SIEDSCHLAG, Carlos; PORTUGAL, Rodrigo. Integrated E&P platform – optimizing data usage within the organization. In: INTERNATIONAL CONGRESS OF THE BRAZILIAN GEOPHYSICAL SOCIETY – SBGf, 18., 2023, Rio de Janeiro.

Al-Rbeawi, S. (2023). A review of modern approaches of digitalization in oil and gas industry. Upstream Oil and Gas Technology, 11. ISSN 2666-2604.

Ahmad, R. W., Salah, K., Jayaraman, R., Yaqoob, I., & Omar, M. (2022). Blockchain in oil and gas industry: Applications, challenges, and future trends. Technology in Society, 68, 101941.

Kang, J., Al Masry, Z., Varnier, C., Mosallam, A., & Zerhouni, N. (2023). Data management framework for risk estimate of electronic boards in drilling and measurement tools. IFAC-PapersOnLine, 56(2), 11936-11941.