

Expert-Guided Axiomatic AI Protocol for Facilitating Subsurface Conceptual Modeling

Jose Dario Rodriguez¹, Kris Piessens¹, Kris Welkenhuysen¹, and Thomas Hermans²

1. Geological Survey of Belgium, Royal Belgian Institute of Natural Sciences (RBINS), Brussels, Belgium; 2. Ghent University (UGent), Ghent, Belgium

The initialization of potential-field implicit geological models and dynamic hydrogeological simulations relies heavily on the rigorous parameterization of spatial constraints (Wellmann & Caumon, 2018).

However, translating qualitative, unstructured legacy data, such as drill logs and historical stratigraphic reports, into quantitative, topologically valid 3D inputs remains a significant industry bottleneck.

Conventional Natural Language Processing (NLP) techniques frequently fail to capture the complex spatial reasoning required for layered subsurface modeling. Recent advances have demonstrated that LLMs, when specialized for domain-specific scientific tasks, can substantially outperform general-purpose NLP approaches in knowledge extraction and reasoning within technical fields (Lin et al., 2024). To bridge this semantic-geometric gap, this study introduces an integrated computational pipeline that couples LLMs with a rigid formal ontology, the Lithotectonic Framework (LTF), a structured rule-based schema developed as part of this work, to support and streamline conceptual model generation.

Rather than treating the LLM as a mere text parser, our protocol deploys it as a structured inference engine constrained by formal ontological rules. The workflow operates in three distinct phases. First, during semantic mapping, the LLM dynamically extracts depth intervals, lithological signatures, and hydrogeological properties from raw textual datasets. Second, the system applies LTF axioms, which prioritize boundaries as the defining features of geological units. By constraining the LLM's output to these strict rules, including constraint-based output filtering and schema validation, the system enforces mathematical and topological consistency, substantially reducing probabilistic inference errors and resolving complex stratigraphic ambiguities. Third, before reaching the human expert, the LLM generates a draft layered conceptual framework that undergoes an internal automated audit via structured self-critique.

The draft is then presented to the geologist through a Human-in-the-Loop (HITL) transparency dashboard. This interface tracks metric uncertainties, data provenance, and underlying assumptions, allowing the expert to actively test and refine the interpretation. This approach substantially reduces the data processing burden associated with legacy datasets while ensuring that final geological interpretations remain under expert human control via defined, auditable protocols. The resulting structured outputs are directly translatable into the coordinate geometries required by 3D interpolation engines. Ultimately, this workflow provides robust foundational data that significantly accelerates the development of layered conceptual models, streamlining their preparation as inputs for complex groundwater flow, geothermal energy, and subsurface storage (CO₂, hydrogen, and natural gas) simulations (Kaufmann & Martin, 2008).

Lin, Z., et al. (2024). "GeoGalactica: A Scientific Large Language Model in Geoscience." arXiv preprint.

<https://doi.org/10.48550/arXiv.2401.00434>

Wellmann, F., & Caumon, G. (2018). "3-D Structural geological models: Concepts, methods, and uncertainties." *Advances in Geophysics*, 59, 1–121. <https://doi.org/10.1016/bs.agph.2018.09.001>

Kaufmann, O., & Martin, T. (2008). "3D geological modelling from boreholes, cross-sections and geological maps, application over former natural gas storages in coal mines." *Computers & Geosciences*, 34(3), 278–290. <https://doi.org/10.1016/j.cageo.2007.09.005>