A Neuro-symbolic knowledge graph framework for geological data integration and interoperability

Emna LAKANI,*

Production Engineering Laboratory (LGP), University of Toulouse, National School of Engineering in Tarbes (ENIT), UTTOP, France

Abstract. Geological data is inherently heterogeneous, fragmented across formats, disciplines, and interpretive frameworks. This research aims to address the challenges of integrating and reasoning over such data by leveraging neuro-symbolic AI and knowledge graph (KG) technologies. We propose a modular KG framework capable of capturing both raw geological observations and layered expert interpretations, while maintaining interoperability and semantic coherence. By combining formal ontologies, symbolic rules, and the extraction capabilities of Large Language Models (LLMs), the approach supports dynamic KG population, interpretation formalization, and multimodal knowledge integration. The neuro-symbolic paradigm further enables reasoning that is both scalable and context-aware, where probabilistic evidence from LLMs is reconciled with logical and domain-specific constraints. This framework is expected to improve knowledge traceability, support domain expert workflows, and promote reuse of geological knowledge across contexts. The final objective is to offer a prototype that empowers geologists to structure, query, and evolve geoscientific knowledge in a unified, machine-interpretable environment.

1 Introduction

Geology is a branch of natural science concerned with the Earth and other astronomical objects, the rocks of which they are composed, and the processes by which they change over time [12]. Geology describes the structure of the Earth, its surfaces and the processes that have shaped that structure. Geology determines the relative ages of rocks found at a given location and geochemistry (a branch of geology) determines their absolute ages. In fact, the present observable state of a rock results from the addition of successive incremental changes acquired during a long and complex history, from their genesis [15]. The geologist's aim is to reconstruct rock history by identifying successive events that made the rocks what they are today. For this reason, he needs to read into various data sources such as geological features description (boreholes, samples, outcrop), structural reference, domain reference, lithostratigraphy reference and geological events reference [20]. Unfortunately, those data are stored in fragmented, structured or semi-structured formats. This fragmentation impedes the integration and reuse of data limiting the generation of actionable insights for interpretation-making. The main challenge remains to establish the links between all the data towards geological event reference. Additionally, a particularly pressing challenge lies in making visible and navigable the web of interpretations that link these diverse data and perspectives. Interpretations are rarely isolated; rather, they are layered, revised, and cross, referenced through time and by different experts. By Far, compounding this technical complexity is the inherently interpretative nature of geology itself. Geological interpretation is not merely a mechanical act of observation, but an epistemic process influenced by the geologist's conceptual frameworks, disciplinary background, and contextual understanding [19][13]. As such, the classification and naming of geological entities often diverge across experts and institutional contexts, leading to variations in conceptual models and domain vocabularies. These divergences introduce semantic inconsistencies and interpretive subjectivity that are difficult to reconcile when attempting to build unified or interoperable geoscientific knowledge systems. Knowledge Graphs (KGs) are graph-based structures that integrate heterogeneous data, capture domain knowledge [10][6], and enable explainable AI through symbolic reasoning. In Geology, where elaborate rock history relies on synthesizing different information (geological observations, lithological observations, structural observations, lithostratigraphic, structural and event reference), we believe that KGs offer the better way to harmonize both raw observations and interpretative reasoning in a structured, machine, interpretable format [14][16]. These technologies provide a robust framework for representing geological entities and their relationships, capturing the nuance of expert judgment, and enabling automated reasoning across distributed data sources. This research proposal examines the challenges and research opportunities in integrating KGs with neuro-symbolic AI, highlighting their potential to enhance explainability, scalability, and modular reasoning to deal with the geological context. The main goal of this work is to develop a prototype for the geologists that not only structure and store data but also empowering knowledge capitalization. The importance of the KG is rooted in the fact that both geological knowledge and data should be integrated to the digital maps.

2 Related works

Geologic semantic ressources

A variety of semantic resources have been developed to support knowledge representation and interoperability in the geosciences. At the general level, GeoSciML [10] has served as a widely adopted conceptual data model for exchanging geological information, such as lithologies, structures, and stratigraphic units. Meanwhile, formal ontologies like SWEET [18] and more recent initiatives such as the Geoscience Ontology (GSO) [5] and GeoCore [9] aim to provide logically grounded vocabularies for Earth science concepts, aligned

* Corresponding Author. Email: emna.lakani@doctorant.uttop.fr

with upper ontologies such as BFO [3]. These foundational models form the semantic backbone for cross-disciplinary data integration.

In more specialized areas, targeted ontologies have emerged to address domain-specific needs. The RESCUE ontology [1] focuses on reservoir characterization in petroleum geoscience, capturing domain-specific concepts such as lithofacies, porosity, and structural traps within a semantic framework that supports reasoning and interpretation. In the field of stratigraphy,[21] developed a detailed ontology covering lithostratigraphic, chronostratigraphic, and biostratigraphic classifications, offering fine-grained semantic modeling of geological units and their relationships. Similar efforts in structural geology include GeoFault [17], which models faults and deformation structures with logical semantics.

Despite their individual strengths, most current ontologies focus on descriptive classification, capturing what geological entities are, rather than how geologists interpret them or how those interpretations evolve. In addition, semantic alignment across subdomains remains limited, with little interoperability between models developed for stratigraphy, petrology, tectonics, and geological time. This fragmentation reveals a critical gap: the absence of an integrated semantic layer capable of linking multiple ontologies while supporting interpretive reasoning and contextual traceability. Addressing this gap is central to the contribution of this research.

Neuro-symbolic approache for dynamic KG construction

The integration of symbolic reasoning and neural models has become a major theme in recent AI research. A recent systematic review of Neuro-Symbolic AI (NSAI) projects from 2020–2024 highlights that most contributions focus on learning and inference, logic and reasoning, and knowledge representation, while important aspects such as explainability and trustworthiness remain less explored [8]. This confirms that the neuro-symbolic paradigm is increasingly seen as a foundation for developing intelligent, context-aware systems that combine the scalability of neural methods with the rigor of symbolic approaches.

Within this landscape, dynamic knowledge graphs (KGs) have received particular attention. Traditional KGs are static, but real-world applications require evolving structures that capture new facts, entities, and relations over time. Recent studies provide formal definitions of dynamic KGs and survey neuro-symbolic methods for tasks such as temporal KG completion and entity alignment [2]. Similarly, dynamic reasoning approaches have been explored in commonsense domains, where knowledge must be generated on demand rather than retrieved from pre-existing graphs. For example, dynamically constructed neuro-symbolic KGs were shown to improve zero-shot commonsense question answering by generating symbolic structures guided by neural models and using them for inference [4]. These works demonstrate how combining probabilistic extraction with symbolic structures enables reasoning that is both flexible and interpretable.

Applications in geosciences are still emerging, but recent advances illustrate the potential of NSAI for scientific domains. A 2025 study on geochemical prediction in copper deposits integrates LLMs, knowledge graphs, and symbolic rules to guide machine learning models. The approach not only improved predictive accuracy but also enhanced interpretability by grounding computational predictions in domain-specific rules and expert knowledge [7]. This confirms the suitability of neuro-symbolic techniques for contexts where uncertainty, interpretive knowledge, and domain-specific constraints are tightly coupled.

Building on these advances, we hypothesize that a dynamic neuro-symbolic KG can provide a suitable reasoning environment for geology. Large language models (LLMs) can be leveraged to extract entities, relations, and interpretive statements from heterogeneous geological sources, while ontologies and expert rules ensure semantic validity and logical coherence. Rather than binary validation, rules and constraints interact with probabilistic evidence to support weighted inference and the emergence of new, context-aware interpretations. Such a system would allow geological knowledge graphs to capture not only raw observations but also evolving interpretations, maintaining traceability and semantic coherence across subdomains

3 Research Questions and key research challenges

According to those facts, we are leading to the following researches questions :

RQ1: *what is the role of KG in order to manage and integrate heterogeneous geological sides ?*

H1: To deal with the geological landscape, it is important to have a unified and logical structured. KGs serves as universal data model, offering a conceptual representation enabling interoperability based on linked data principles. the main idea is to integrate existing semantic resources to establish this graph in a modular way. The KG is partitioned according to geologic specific features.

RQ2: *Can a modular knowledge graph architecture, driven by query context, support scalable and domain specific reasoning in geologic applications?*

H2: A modular, query-driven KG system improves computational efficiency and semantic relevance by dynamically loading only the required ontology modules.

RQ3: How can neuro-symbolic approaches, combining Large Language Models (LLMs) with ontologies and domain-specific rules, support dynamic knowledge graph construction and probabilistic reasoning in geology?

H3: LLMs can assist in extracting candidate entities, relations, and geological processes, while ontologies and expert rules provide formal and domain-grounded validation. Their integration in a neuro-symbolic setting enables the construction of dynamic knowledge graphs, where probabilistic evidence and symbolic constraints interact to support coherent, traceable, and evolving geological interpretations.

4 Conclusion and future Work

This research, still in its early stage, proposes a neuro-symbolic knowledge graph approach to address the integration and reasoning challenges of geological data. At present, the work focuses on a survey of existing semantic resources in geology to evaluate their coverage and identify gaps. The next step will be to build a geological ontology by aligning these resources and extending them with missing modules for domains. This ontology will serve as the backbone of a modular KG that integrates heterogeneous sources and captures both observations and interpretations. LLMs will then be employed to extract and populate knowledge from geological corpora, while symbolic rules and logical constraints ensure semantic validity and coherence. The framework will be progressively tested and refined with the involvement of geology experts to evaluate its interpretability and usability. The long-term goal is to deliver a dynamic and expert-validated system that supports geological interpretation, strengthens knowledge traceability, and enhances the integration of geoscientific knowledge into digital maps.

5 Acknowledgments

This work is part of digital earth platform witch is supported by PEPER sous-sol project ANR-22-EXSS-0005.

References

[1] M. Abel, S. Gérard, C. Jean, and F. Bosquet. The rescue ontology: Modeling geological reservoirs for petroleum exploration. *Computers & Geosciences*, 35(8):1613–1624, 2009.

[2] M. Alam, G. A. Gesese, and P.-H. Paris. Neurosymbolic methods for dynamic knowledge graphs. In *Dynamic Knowledge Graphs*, Studies on the Semantic Web. IOS Press, Amsterdam, Netherlands, 2024. Chapter contribution.

[3] R. Arp, B. Smith, and A. D. Spear. *Building Ontologies with Basic Formal Ontology*. MIT Press, 2015.

[4] A. Bosselut, R. L. Bras, and Y. Choi. Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2328–2339, Online, 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.182. URL https://doi.org/10.18653/v1/2020.emnlp-main.182.

[5] B. Brodaric and S. M. Richard. The geoscience ontology (gso). Geological Survey of Canada and USGIN Foundation Inc., 2021.

[6] T. Callahan, I. Tripodi, A. Stefanski, L. Cappelletti, S. Taneja, J. Wyrwa, E. Casiraghi, N. Matentzoglu, J. Reese, J. Silverstein, C. Hoyt, R. Boyce, S. Malec, D. Unni, M. Joachimiak, P. Robinson, C. Mungall, E. Cavalleri, T. Fontana, G. Valentini, M. Mesiti, L. Gillenwater, B. Santangelo, N. Vasilevsky, R. Hoehndorf, T. Bennett, P. Ryan, G. Hripcsak, M. Kahn, M. Bada, and L. Hunter. An open-source knowledge graph ecosystem for the life sciences. https://arxiv.org/abs/2307.05727, 2024.

[7] W. Chen, J. Zhang, W. Li, X. Que, C. Li, and X. Ma. Integrating neuro-symbolic ai and knowledge graph for enhanced geochemical prediction in copper deposits. *Applied Computing and Geosciences*, 27:100259, 2025. doi: 10.1016/j.acags.2025.100259.

[8] B. C. Colelough and W. Regli. Neuro-symbolic ai in 2024: A systematic review. In *Logical Foundations of Neuro-Symbolic AI 2024 (co-located with IJCAI 2024)*, volume 3819 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2025. ISBN 1613-0073. Masters Along the Way scholarly paper; presented at the Logical Foundations of Neuro-Symbolic AI 2024 workshop, Jeju Island, South Korea, August 5 2024; also available as arXiv:2501.05435.

[9] L. F. Garcia, M. Abel, M. Perrin, and R. d. S. Alvarenga. The geocore ontology: A core ontology for general use in geology. *Computers & Geosciences*, 177:105305, 2023. doi: 10.1016/j.cageo.2023.105305.

[10] G. S. W. Group. Ogc geosciml: A model for geological information interchange (v4.0). Open Geospatial Consortium Standard, 2015.

[11] C. Gutiérrez and J. F. Sequeda. Knowledge graphs. *Communications of the ACM*. doi: 10.1145/3418294. Review Article.

[12] D. Harper. Geology. https://www.etymonline.com/word/geology, n.d. Online Etymology Dictionary.

[13] B. Le Bayon, M. Padel, T. Baudin, F. Cagnard, B. Issautier, H. Tissoux, C. Prognon, A. Plunder, S. Grataloup, F. Lacquement, A. Hertout, and J. Stephan-Perrey. The geological-event reference system, a step towards geological data harmonization. *BSGF - Earth Sciences Bulletin*, 2022. doi: 10.1051/bsgf/2022017. URL https://doi.org/10.1051/bsgf/2022017. Journal Article.

[14] H. Li, G. Appleby, C. Brumar, R. Chang, and A. Suh. Knowledge graphs in practice: Characterizing their users, challenges, and visualization opportunities. *IEEE Transactions on Visualization and Computer Graphics*. doi: 10.1109/TVCG.2023.3326904.

[15] H. G. Liddell and R. Scott. A Greek–English Lexicon, Perseus Project.

[16] A. Polleres, R. Pernisch, A. Bonifati, D. Dell'Aglio, D. Dobriy, S. Dumbrava, L. Etcheverry, N. Ferranti, K. Hose, E. Jiménez-Ruiz, M. Lissandrini, A. Scherp, R. Tommasini, and J. Wachs. How does knowledge evolve in open knowledge graphs? *Transactions on Graph Data and Knowledge*, 2023. doi: 10.4230/TGDK.1.1.11.

[17] L. Qui, J. Chen, and H. Wang. Geofault: An ontology for modeling faults and structural geology features. *Geoscience Frontiers*, 14(1):101–112, 2023.

[18] R. Raskin and M. J. Pan. Semantic web for earth and environmental terminology (sweet). In *Workshop on Earth Observation and the Semantic Web*, 2005.

[19] Skrzypek. *Contribution structurale, pétrologique et géochronologique à la tectonique intracontinentale de la chaîne hercynienne d'Europe (Sudètes, Vosges)*. PhD thesis, Université de Strasbourg, 2012. PhD Thesis, Earth Sciences.

[20] The Geological Society. What is geology?, 2023. Accessed May 31, 2023.

[21] Y. Xu, L. Zhang, S. Liu, et al. A stratigraphic ontology for semantic representation of geological time and units. *Earth Science Informatics*, 15:287–305, 2022.