

# Challenges and Approach: AI for Digitalised Carbon Storage Analysis

Yuanwei Qu<sup>1</sup>, Arild Waaler<sup>1</sup>, Anita Torabi<sup>2</sup> and Baifan Zhou<sup>3,1</sup>

<sup>1</sup>Department of Informatics, University of Oslo, Norway

<sup>2</sup>Department of Geosciences, University of Oslo, Norway

<sup>3</sup>Department of Computer Science, Oslo Metropolitan University, Norway

## Abstract

Carbon Capture and Storage (CCS) plays an essential role in mitigating climate change. A comprehensive CO<sub>2</sub> storage analysis is key to CCS success, however, its advancement is hindered by interdisciplinary complexities, data fragmentation, and the need for enhanced transparency. This paper presents ongoing work in an integrated digitalisation framework for CO<sub>2</sub> storage analysis that addresses three major challenges: knowledge complexity across geosciences, reservoir engineering, and computer science; the heterogeneous, multi-scale nature of geological data; and the requirement for explainable decision-making in high-risk subsurface storage scenarios. Our approach leverages both symbolic AI and generative AI. On the symbolic AI side, it includes ontology-driven knowledge engineering to harmonise domain-specific terminologies, and employs a layered information modelling and data integration system to standardise diverse datasets. Additionally, a generative AI-powered query and explanation system provides context-aware, transparent analyses. This framework aims to streamline multidisciplinary collaboration and enhance the accountability and reliability of long-term CO<sub>2</sub> storage, thereby supporting the transition to a carbon-neutral future.

## Keywords

carbon capture and storage, knowledge engineering, information modelling, explainable AI

## 1. Introduction

Carbon Capture and Storage (CCS) captures CO<sub>2</sub> from industrial sources or the atmosphere and securely stores it in geological formations. This technology mitigates greenhouse gas emissions and supports the transition to net-zero emissions [1]. To achieve its full potential in global CO<sub>2</sub> mitigation and meet predicted storage capacities, CCS requires advanced, cost-effective analysis methods that ensure long-term storage safety and integrity [2, 3]. A key advancement in this area is digital transformation: transitioning from traditional workflows to data- and knowledge-driven methods that improve the efficiency and accuracy of storage analysis.

However, digitalising carbon storage analysis remains challenging despite its recognised importance in the energy sector [4]. This challenge arises from the need for CO<sub>2</sub> storage analysts to integrate vast amounts of domain-specific knowledge and analyse data in various formats, from structured datasets to images and diagrams. While machine learning advances offer promise for specific tasks [5], their accuracy is often constrained by the unstructured nature of CO<sub>2</sub> storage data. While machine learning advances offer promise for specific tasks [5], their effectiveness is often constrained by the unstructured nature of CO<sub>2</sub> storage data due to accessibility, quality, and heterogeneity issues.

Compounding these technical difficulties, the interdisciplinary nature of CO<sub>2</sub> storage analysis requires collaboration across domains such as geoscience, reservoir engineering, and computer science. Each domain has its unique methodologies and terminologies, hindering effective collaboration due to differing approaches and tools, complicating data processing and interpretation. In this context, we characterise three major challenges for the design and operation of digitalised carbon storage analysis:

---

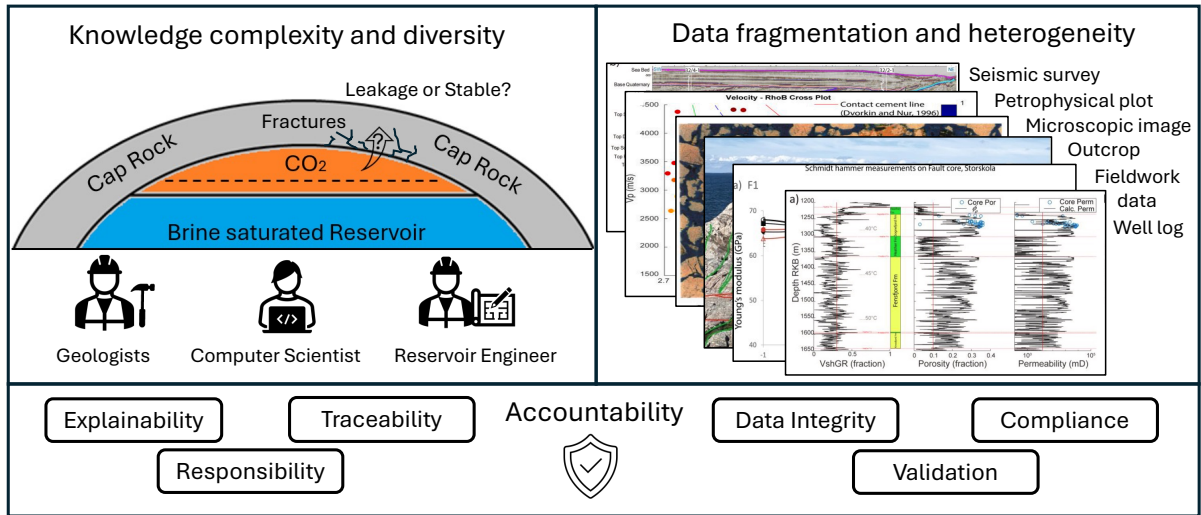
NAIS 2025: Symposium of the Norwegian AI Society, June 17–18, 2025, Tromsø, Norway

✉ quyu@ifi.uio.no (Y. Qu); arid@ifi.uio.no (A. Waaler); anita.torabi@geo.uio.no (A. Torabi); baifan.zhou@oslomet.no (B. Zhou)

ORCID 0000-0003-3220-2101 (Y. Qu); 0000-0002-6228-1317 (A. Waaler); 0000-0001-5685-808X (A. Torabi); 0000-0003-3698-0541 (B. Zhou)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



**Figure 1:** Multidimensional challenges in cap rock analysis in CO<sub>2</sub> storage : (top left) A fractured cap rock and CO<sub>2</sub> reservoir, experts from different domain might have different views; (top right) Examples of required heterogeneous data for analysis (modified from [6, 7]); (bottom) Accountability through explainability, traceability, responsibility, and data integrity, compliance and validation.

- C1, *Knowledge complexity and diversity* highlights the need to reconcile disparate terminologies and conceptual frameworks across geoscience, reservoir engineering, and computer science. Experts may struggle to interpret concepts or data presented in another discipline's terminology.
- C2, *Data fragmentation and heterogeneity* is a critical issue, as the CO<sub>2</sub> storage analysis relies on heterogeneous data types such as well logs, seismic surveys, core samples and analogues from the onshore fields. These data sources are highly fragmented cross data silos, rendering manual data interpretation and integration by human experts inefficient.
- C3, *Accountability* is crucial, as CO<sub>2</sub> storage processes are high-stake. All aspects of analysis and system operation, including data queries and knowledge retrieval, must be transparent and easily explainable to ensure trust, accountability, and reliability.

To address these challenges, we propose an integrated approach. For C1, *knowledge engineering* bridges semantic gaps across disciplines to create a *unified conceptual framework*. C2 is addressed through *information modelling and data integration* techniques that standardise and process diverse data sets, providing *unified data access*. Finally, C3 is tackled with an *AI-enhanced system for data and knowledge query* that provides transparent, context-aware responses and explanations for storage analyses and knowledge queries. Together, these strategies aim to enable effective interdisciplinary collaboration, supported by high-quality data and knowledge, positioning digitalised carbon storage analysis as a key element in decarbonisation efforts.

## 2. Challenges

Digitalising carbon storage analysis demands tight collaboration among geoscientists, reservoir engineers, and computer scientists. This section elaborates on the three major challenges, with examples illustrated in Figure 1.

### 2.1. Challenge C1: Knowledge Complexity and Diversity

CO<sub>2</sub> storage analysis spans multiple disciplines, each with its own specialised knowledge and conceptual models. These frameworks are not only complex individually but also diverse in terminology and perspective. For instance, the term *fracture* can vary in meaning: reservoir engineers may describe it as lines or 2D planes in models, while geoscientists may define it as a 3D volume network with a core and damage zone. These semantic differences affect the interpretation of caprock integrity analysis, potentially lead-

ing to inconsistent leakage predictions. Computer scientists, unfamiliar with these domain nuances, face difficulties understanding such contextual variations. Furthermore, existing analysis methods are largely adapted from petroleum engineering, which prioritises extraction. In contrast, CO<sub>2</sub> storage demands a focus on long-term containment, requiring a shift in both objectives and conceptual understanding.

## 2.2. Challenge C2: Data Fragmentation and Heterogeneity

CO<sub>2</sub> storage analysis depends on diverse geological datasets that are often fragmented across isolated systems. Data may originate from well logs that provide detailed measurements of subsurface properties, seismic surveys that map geological structures, core samples that offer physical insights into rock composition, and onshore outcrop data that provide analogue scenarios for geologists to reference and build geo-mechanical models to predict the behaviour of reservoir and cap-rock. These datasets differ not only in format and resolution but also in the type of information they convey. For example, seismic data are often high-dimensional and require sophisticated signal processing, while well log data are typically time-series measurements with varying scales. This fragmentation and heterogeneity hinder effective data access and analysis, complicating data interpretation and limiting the efficiency of both human analysis and data science methods.

## 2.3. Challenge C3: Accountability

CO<sub>2</sub> storage involves high-stake infrastructure where decisions directly affect safety, environmental impact, and operational efficiency. As AI-supported methods become increasingly integral to storage analysis [8], ensuring transparency, explainability, and traceability in decision-making is essential. For instance, during the CO<sub>2</sub> injection phase, when an anomaly—such as a sudden pressure change in a storage reservoir is detected, the geologists and reservoir engineers need to understand the underlying reasoning behind any automated response or control adjustment. Accountability also demands that data queries and knowledge retrieval are traceable and interpretable, supporting compliance with regulations, verification of data integrity, and clear responsibility in operations.

Each of these challenges highlights the complexity inherent in creating a robust digitalised CO<sub>2</sub> storage analysis method. In the following section, we describe our approach that targets these challenges through a combination of knowledge engineering, information modelling with data integration, and large language model-based query and explanation systems.

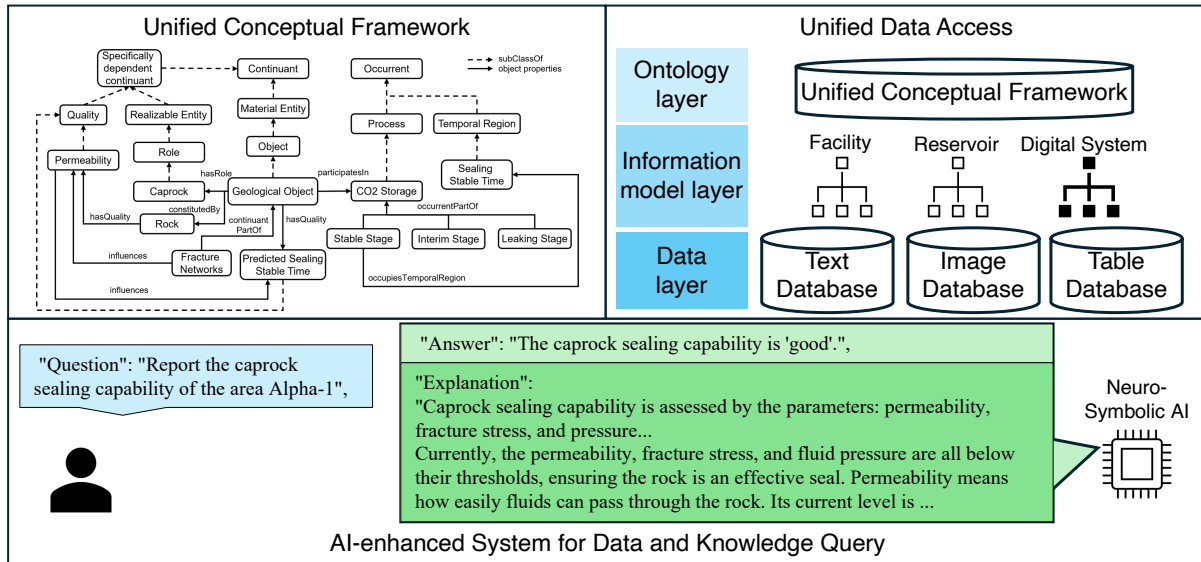
# 3. Approach

To overcome the challenges in digitalising carbon storage analysis, our approach is structured around three key strategies, each targeting a specific challenge (Figure 2).

## 3.1. Knowledge Engineering

To address knowledge complexity and diversity (C1), we establish a *unified conceptual framework* that aims to bridge the semantic gaps across geoscience, reservoir engineering, and computer science. Our approach employs knowledge engineering techniques to develop a unified ontology that harmonises domain-specific terminologies and concepts. For example, the ontology should define terms such as *cap rock*, *sealing stable time*, *permeability*, etc., with machine-readable definitions that are agreed upon by domain experts. In the current knowledge model (Unified Conceptual Framework in Figure 2), we separate the concept of *cap rock* from rock as material and make a conceptual distinction between the temporal region of the *sealing stable time* and the predicted value of storage time.

To not reinvent the wheel, this framework is built above the existing useful ontological resources such as top-level ontology: Basic Formal Ontology [9], core ontology: GeoCore Ontology [10], and domain-specific ontology: GeoFault Ontology [11]. By mapping and clarifying these concepts into a unified conceptual model, this framework enables experts from different domains to share and interpret information accurately, supporting effective multidisciplinary collaboration.



**Figure 2:** Our approach to digitalised CO<sub>2</sub> storage analysis: a unified conceptual framework (knowledge engineering) harmonises domain-specific terms and concepts; a unified data access framework (information modelling and integration) standardises access to diverse geological and operational data; and an AI-enhanced query system supports transparency and accountability.

### 3.2. Information Modelling and Data Integration

The second challenge (C2) relates to the fragmented multi-scale data sources, such as petrographic thin section, well logs, core samples, seismic surveys, and onshore outcrops. Our approach aims at a *unified data access* framework, enabling a consistent and structured way of accessing diverse data through a common semantic and modelling framework, comprising of three layers:

- The ontology layer that relies on the unified conceptual framework to ensure clarity and consistent terminology.
- The information model layer that serves as a mediator between the ontology layer and data layer by constructing information models from the perspectives of geoscience, reservoir engineering, and computer science, and by implementing data integrity and constraint checks.
- The data layer are mapped to and accessed via the information model layer, comprising of diverse data, such as text data (e.g., documents, reports), image data (e.g., tomography, seismic survey), table data (e.g., sensor data).

The *unified data access* approach organises geological and operational data into a structured, query-able format that can be efficiently accessed and interpreted by both humans and automated systems. To ensure the work results are well aligned with industrial standards, the construction of *information model* will follow the recommended practice for asset information modelling framework from DNV [12] and apply the suggested Information Modelling Framework language [13].

### 3.3. AI-enhanced System for Data and Knowledge Query

Given the risks associated with CO<sub>2</sub> storage analysis and the followed operating processes, accountability, including transparency and explainability (C3) are critical. To meet this need, we aim to develop an *AI-enhanced system for data and knowledge query* that provides context-aware, human-understandable answers and explanations. The system relies on the *unified data access* and provides semantic-enriched responses. The system employs retrieval-augmented generation (RAG) techniques to draw upon up-to-date domain knowledge and operational data, ensuring that each decision or output is accompanied by clear and traceable explanations. For example, when an operator queries the caprock sealing capability, which is crucial for ensuring that the subsurface carbon storage system does not leak, the system will not only assess the sealing capability based on the retrieved relevant sensor data, but also explain the

knowledge and reasoning behind the assessment in plain language (giving explanations). This approach enhances operators' trust and supports informed decision-making in complex subsurface environments.

## 4. Summary

This paper presents our ongoing effort to develop a data- and knowledge-driven approach enhanced by AI for CO<sub>2</sub> storage analysis. We identify three core challenges: the complexity and diversity of domain-specific knowledge, the heterogeneous and multi-scale nature of data from different sources, and the critical need for transparency and explainability raised by accountability in carbon storage analysis and operations. To address these challenges, our approach integrates: (1) knowledge engineering to create a unified conceptual framework, (2) information modelling and data integration to standardise and fuse diverse datasets, enabling unified data access aligned with industry best practices, and (3) a generative AI-enhanced system for data and knowledge query to support accountable, transparent, and traceable decision-making. Together, these components form a foundation for a more efficient, reliable, collaborative, and digitalised CO<sub>2</sub> storage analysis framework, contributing to a carbon-neutral future.

**Declaration on Generative AI.** The authors have used ChatGPT to assist with the polishing of human-authored text. The authors take full responsibility for the publication's content.

## References

- [1] IEA, Net zero by 2050, 2021. URL: <https://www.iea.org/reports/net-zero-by-2050>, accessed 10 April 2025.
- [2] T. Kazlou, A. Cherp, J. Jewell, Feasible deployment of carbon capture and storage and the requirements of climate targets, *Nature Climate Change* 14 (2024) 1047–1055.
- [3] R. Gholami, A. Raza, S. Iglauer, Leakage risk assessment of a co2 storage site: A review, *Earth-Science Reviews* 223 (2021) 103849.
- [4] IEA, Digitalisation and energy, 2017. URL: <https://www.iea.org/reports/digitalisation-and-energy>, accessed 9 April 2025.
- [5] F. Hussin, S. A. N. M. Rahim, N. S. M. Hatta, M. K. Aroua, S. A. Mazari, A systematic review of machine learning approaches in carbon capture applications, *Journal of CO2 Utilization* 71 (2023).
- [6] M. Fawad, M. J. Rahman, N. H. Mondol, Seismic reservoir characterization of potential co2 storage reservoir sandstones in smeaeheia area, northern north sea, *Journal of Petroleum Science and Engineering* 205 (2021).
- [7] A. Torabi, B. Alaei, A. Smith, Fault characteristics in exhumed basement rocks; implications for understanding subsurface basement faults, *Tectonophysics* 887 (2024) 230445.
- [8] M. Nassabeh, Z. You, A. Keshavarz, S. Iglauer, Sub-surface geospatial intelligence in carbon capture, utilization and storage: a machine learning approach for offshore storage site selection, *Energy* 305 (2024).
- [9] R. Arp, B. Smith, A. D. Spear, Building ontologies with basic formal ontology, Mit Press, 2015.
- [10] L. F. Garcia, M. Abel, M. Perrin, R. dos Santos Alvarenga, The geocore ontology: a core ontology for general use in geology, *Computers & Geosciences* 135 (2020) 104387.
- [11] Y. Qu, M. Perrin, A. Torabi, M. Abel, M. Giese, Geofault: A well-founded fault ontology for interoperability in geological modeling, *Computers & Geosciences* 182 (2024) 105478.
- [12] DNV, DNV-RP-0670, Recommended practice for asset information modelling framework, 2024. URL: <https://www.dnv.com/digital-trust/recommended-practices/asset-information-modelling-dnv-rp-0670/>, accessed 10 April 2025.
- [13] A. Waaler, M. Skjæveland, (editors), Information modelling framework manual version 0.3.0, 2024. URL: <https://www.imfid.org/>, accessed 10 April 2025.