Quarterly Report – Public Page

Date of Report: *7th Quarterly Report – June 30, 2025*

Contract Number: *693JK32310007POTA*

Prepared for: *DOT-PHMSA*

Project Title: *An Integrated Knowledge Graph Model for Geohazard Monitoring Data*

Prepared by: *University of Wyoming*

Contact Information: *Dr. Minou Rabiei, mrabiei@uwyo.edu*

For quarterly period ending: *June 30, 2025*

1: Items Completed During this Quarterly Period:

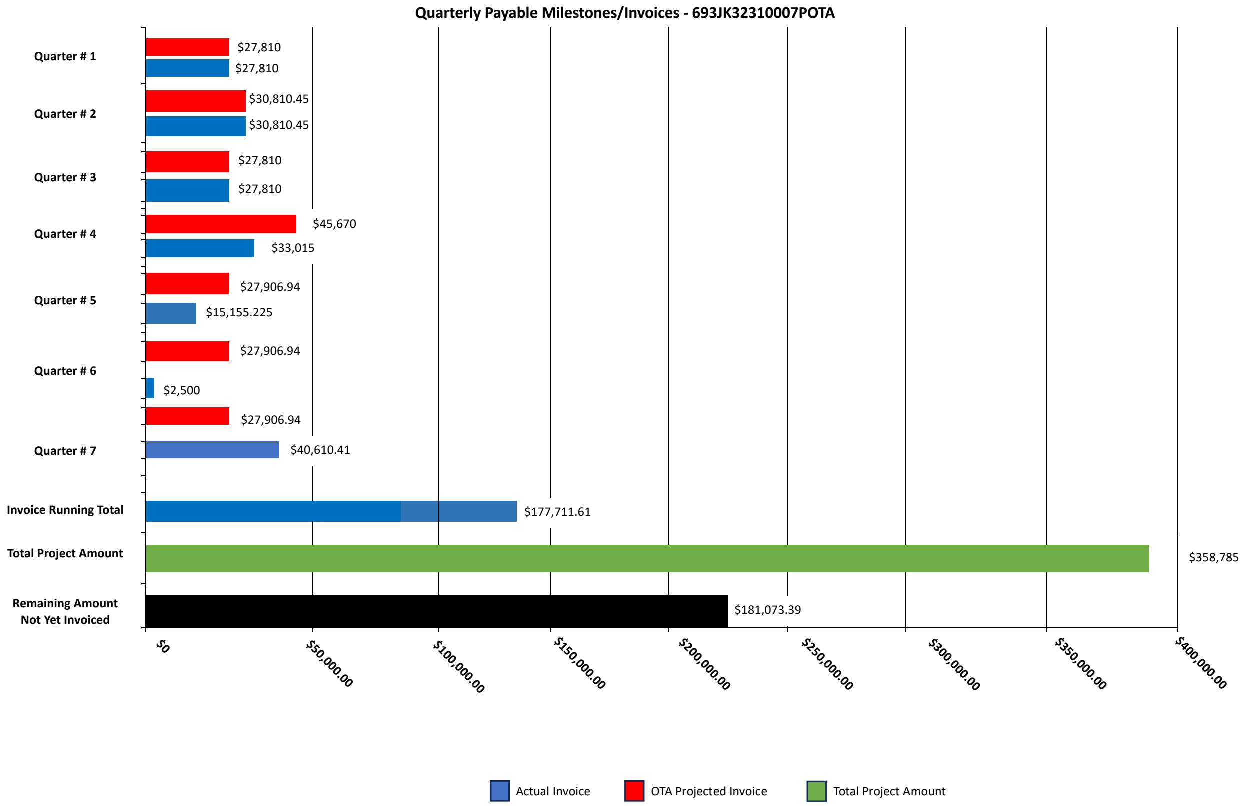
- Mapped the collected and standardized data into the defined ontology
- Integrated data from various sources based on the determined mapping

2: Items Not-Completed During this Quarterly Period:

- Link external data sources and perform semantic enrichment (On-going)
- Complete the design and construction of knowledge graph model
- Develop the graph neural network model

Justification: Due to staffing issues, we experienced delays in the development of the ontology model, mapping of the collected and standardized data into the defined ontology and integrating data from various sources based on the determined mapping. The staffing issue is now resolved, and we have successfully developed the ontology model during this quarter. The project team will continue working on Tasks 20, 22 and 23 during the next quarter.

3: Project Financial Tracking During this Quarterly Period:



4: Project Technical Status

Narrative: The earlier stages of the project were dedicated to systematically ingesting and preprocessing heterogeneous datasets from authoritative sources such as USGS (seismic and landslide events), PHMSA (pipeline incidents), FEMA (flood zones), PRISM (climate normals), and NHD/HUC-12 (hydrology). This phase involved extracting, cleaning, standardizing, and reprojecting spatial and temporal data to ensure consistency and interoperability.

Once these preprocessing tasks were complete, the next step was to design a semantic structure that could support meaningful integration and complex querying across these diverse data types. This led us to the development of a domain-specific ontology model tailored for the geohazard domain.

What is an Ontology Model?

In the context of semantic web technologies, an ontology is a formal specification of concepts (classes), their properties (relationships), and constraints that describe a particular domain of knowledge. It defines:

- **Classes** (e.g., Pipeline, EarthquakeHazard, Watershed),
- **Properties** (e.g., occursInState, hasGeometry, hasDateTime),
- **Hierarchies and relationships** (e.g., SeismicEvent is a subclass of HazardEvent),
- **Domains and ranges**, which constrain how entities are connected.

Ontologies provide the semantic scaffolding for structuring data in a machine-interpretable way, making it possible to link, query, and reason across otherwise incompatible datasets.

A Knowledge Graph is more than just a linked dataset; it is a semantically rich structure that enables context-aware querying and reasoning. Ontologies are essential for KG development because they:

- **Enable Interoperability**: By defining shared vocabularies and relationships, ontologies allow integration of diverse datasets under a common schema.
- **Support Inference**: Ontologies provide logic-based structures that allow the KG to infer new knowledge from existing facts (e.g., if an earthquake is within a watershed, it may affect downstream flowlines).
- **Enhance Query Capability**: SPARQL queries rely on the ontology to interpret property paths, class hierarchies, and constraints, enabling complex spatiotemporal and domain-specific questions.
- **Facilitate Data Validation**: Ontologies define the expected structure and relationships, helping to detect inconsistencies, missing data, or schema violations during the data loading process.
- **Provide Reusability and Extensibility**: A well-structured ontology can be reused in future projects and extended to accommodate new data types or domains (e.g., climate change impacts, soil stability).

Our ontology model provides a semantic framework for integrating heterogeneous datasets related to geohazards, pipelines, administrative regions, and environmental factors. The design follows modular principles, with domain-specific sub-ontologies that interlink through shared entities and properties. The purpose is to enable advanced querying, reasoning, and analytics within a knowledge graph (KG) environment, supporting applications such as risk assessment, early warning systems, and infrastructure management.

The ontology aligns with well-established vocabularies (e.g., RDF, OWL, GeoSPARQL, TIME ontology) to ensure semantic interoperability and extensibility. The modules are designed to support spatial, temporal, and event-based reasoning, while enabling linkage to administrative and infrastructure data.

Domain Analysis and Vocabulary Selection

The development began with a detailed analysis of the data sources—USGS, PHMSA, FEMA, PRISM, and others. Based on the schema and semantics of these datasets, the following namespaces were adopted:

- **Prefixes**: ex, rdf, rdfs, owl, xsd, geo, time, skos, and schema. These enable integration with GeoSPARQL, temporal reasoning, and standard RDF serialization.

Modular Ontology Design

The modular ontology provides a well-structured semantic foundation for integrating and querying complex, heterogeneous geospatial and infrastructure datasets. The use of spatial and temporal ontologies allows for fine-grained reasoning, while the hazard and pipeline modules support risk analytics and decision support. Designed for extensibility and real-time integration, this ontology is a key enabler of predictive and preventive geohazard management.

After finalizing the modular ontology design, we loaded the integrated ontology into Apache Jena Fuseki, a SPARQL-compliant triple-store that supports GeoSPARQL extensions and reasoning over RDF data. During this process, we ingested a total of approximately 28.7 million RDF triples, which were distributed across eight Turtle (TTL) files, each corresponding to a major thematic data source (e.g., seismic events, fault lines, flood zones, climate normals, flowlines, watersheds, pipeline incidents, and administrative boundaries).

To enable meaningful semantic querying and inferencing across heterogeneous datasets, we asserted a set of core object and datatype properties, particularly those facilitating spatiotemporal integration. These assertions allow for spatiotemporal joins across conceptually different datasets, such as aligning climate patterns with pipeline incidents, or locating seismic events within hydrological units and administrative boundaries. By asserting these relationships explicitly within the knowledge graph, we enable cross-domain, multi-resolution querying; one of the primary advantages of semantic integration. Researchers and stakeholders can now ask complex questions that involve both physical infrastructure and environmental phenomena across time and space.

Current Limitations and Next Steps

Despite promising results in query performance and semantic reasoning, we acknowledge several current limitations:

- **Pipeline Geometry Gap**: Detailed ex:PipelineSegment geometries are currently not available or are only partially defined. Without line geometries and topological connectivity, we cannot yet perform precise spatial joins between pipelines and intersecting hazard zones. Addressing this will be a priority in future data acquisition and ingestion cycles.
- **Predictive Insights**: The current knowledge graph is designed for retrospective analysis and descriptive querying of past events. It does not yet incorporate predictive models or reasoning components (e.g., risk scores, failure probability estimations). We plan to integrate machine learning components and temporal forecasting modules that will augment the KG with forward-looking capabilities, transforming it into a more proactive decision-support system.

While the core ontology has been implemented successfully and foundational SPARQL-based validations have confirmed its operational integrity, this phase marks a transition point from structural development to iterative refinement. The ontology will continue to evolve as new datasets are ingested, new relationships are modeled, and new analytical capabilities are added. This iterative approach ensures that the KG remains relevant, scalable, and analytically powerful for geohazard and infrastructure risk management.