

Geosemantics framework in a Nutshell

- [Abstract](#)
- [Science Challenges](#)
- [Goals and Vision](#)
- [Design Overview](#)
- [Key technologies used in the framework](#)
- [Contribution](#)
 - [Scientific Contribution](#)
 - [Technical contribution](#)

Abstract

Many challenges hinder the seamless integration of models with data. These challenges compel scientists to perform the integration process manually. The primary challenges are a consequence of the knowledge latency between model and data resources and others are derived from inadequate adoption and exploitation of information technologies. Knowledge latency challenges increase exponentially when a user aims to integrate long-tail data (data collected by individual researchers or small research groups) and long-tail models (models developed by individuals or small modeling communities). We focus on these long-tail resources because despite their often-narrow scope, they have significant impacts in scientific studies and present an opportunity for addressing critical gaps through automated integration. The goal of this research is to develop a framework rooted in semantic techniques and approaches to support "long-tail" models and data integration. Our vision is to develop a decentralized knowledge-based platform that can be easily adopted across geoscience communities comprising of individual and small group researchers, to allow semantically heterogeneous system to interact with minimum human intervention. It will allow the automatic reference of data from data resources to model by: (i) leveraging the Semantic Web; (ii) developing an automated semantic mediation tool; and (iii) developing a semantic knowledge discovery system that can be used by long-tail models.

Science Challenges

Our goal is to enable the integration of long-tail data, i.e. data collected by individual researchers or small research groups, and long-tail models, i.e. models developed by individuals or small modeling communities, using a framework rooted in semantic techniques. We focus on these long-tail resources because despite their often-narrow scope, they have significant impacts in scientific studies and present an opportunity for addressing critical gaps through automated integration. We aim is to develop a decentralized knowledge-based platform that can be easily adopted across geoscience communities comprising of individual and small group researchers, to allow semantically heterogeneous system to interact with minimum human intervention.

Goals and Vision

Develop a decentralized knowledge-based platform that allows semantically heterogeneous systems to interact with minimum human intervention.

We will build on two existing technologies:

- SEAD (Sustainable Environmental Actionable Data): it supports the full life-cycle of long-tail data including collection, curation, discovery, sharing, and preservation.
- CSDMS (Community Surface Dynamics Modeling System): it supports the conversion of existing models into a plug and play system for interoperable integration.

We will also integrate with ongoing EarthCube initiatives including GeoSoft, Earth System Bridge, SEN (Sediment Experimentalist Network), and eWELL (Workforce Education and Learning Library).

Design Overview

The framework consists of three layers:

- knowledge base layer: it stores the Standard Names graph and semantically enabled models.
- knowledge management layer: it is responsible for reasoning, logic ingestion, semantic processing, and indexing of new resources.
- Web application layer: it deploys the four web services.

Key technologies used in the framework

- Play framework is used as a web application framework.
- Services are coded in Scala, python, and Java.
- Jena TDB is used store the SN graph.
- Fuseki DB is used to store the triples of the Geosemanitc wiki.

Contribution

Scientific Contribution

Geosemantics framework will directly augment the multidisciplinary interaction between different geoscience communities by minimizing the human intervention in semantic mediation between resources and their context ambiguity, and supporting the "crosswalks" among geoscience Standard Names.

Technical contribution

- Graph knowledge base for storing linked Standard Names.
- GeoSemantic Wiki system for geoscience communities to annotate their Standard Names.
- Knowledge Discovery Service for retrieving the graph of a data node and infer the contextual association between resources.
- Semantically enabled models as a foundation for advancing Model-as-a-Service.
- Resources Alignment Service for handling the semantic mediation between model and data resources.
- Semantic Annotation Service for annotating resources with standard names, `\emph{encapsulated Standard Names}`, and incorporating semantics in the development of models
- Knowledge Integration Service for ingesting Standard Names, reasoning over their definition, and code the `inferred` relationships using SKOS vocabularies.