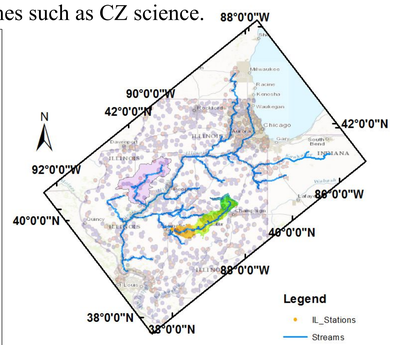
A Geo-Semantic Framework for Integrating Long-Tail Data and Models

Praveen Kumar, Mostafa Elag, Luigi Marini, Peishi Jiang, Rui Liu, and Yao Hu (UIUC)
Scott D. Peckham, and Maria Stoica (CSU), Leslie Hsu, and Kim Miller (LDEO)

Motivation

Integration of multidimensional and heterogeneous data with models is essential in all geoscience disciplines such as CZ science.

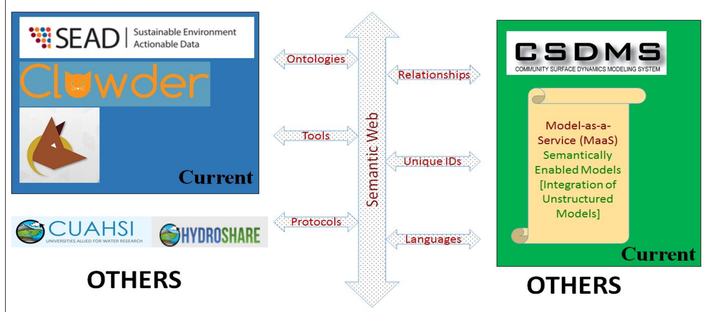
For example, in Illinois River Basin there are more than 9000 environmental and streamflow monitoring stations, which are continuously producing data over various spatial and temporal resolutions. Similarly multiple sophisticated models are required to be seamlessly coupled and integrated with data to support new inter-disciplinary investigations.



Incorporation of semantics in data and models life cycles to support:

- **Data-model seamless integration** by overcoming the semantic heterogeneity of the rapidly growing data and model collections.
- **Semantic data discovery** by identifying the semantic relationship "crosswalks" among standard names
- **Data synthesis** based on their information profile to implement the Linked Data standards.

Technical Approach



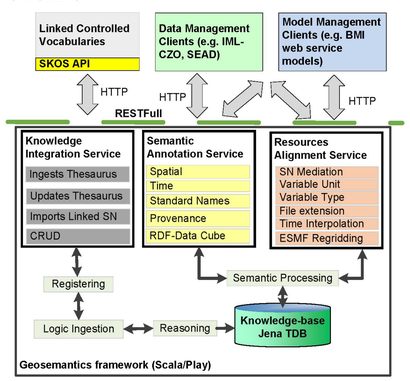
Framework Architecture

Geosemantic framework uses:

- Micro-service architecture
- Linked Data standards.

It provides:

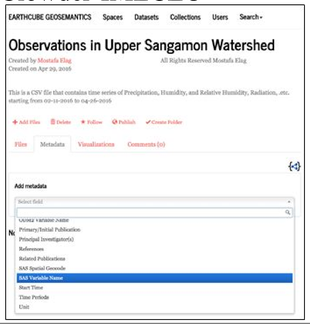
- Semantic Annotation Services (SAS)
- Resources Alignment Services (RAS)
- Knowledge Integration Services (KIS)



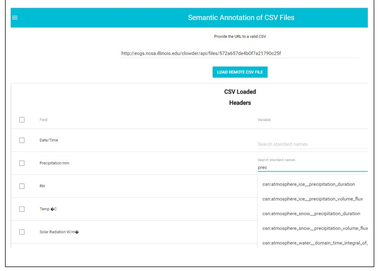
Semantic Workflow for Integration of Long-Tail Data and Model

1- Semantic Annotation Services (SAS)

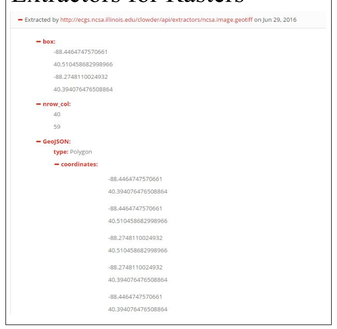
Semantic Annotation in Clowder-IMLCZO



Sensor data annotation (*.CSV)
http://hcgs.ncsa.illinois.edu/csv



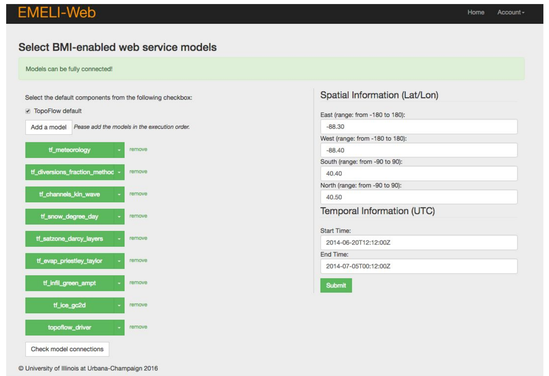
Semantic Information Extractors for Rasters



2- Model coupling environment: EMELI-Web

EMELI-web is a web application for EMELI (Experimental Modeling Environment Framework for Linking and Interoperability)
http://ecgs.ncsa.illinois.edu/emeli-web/

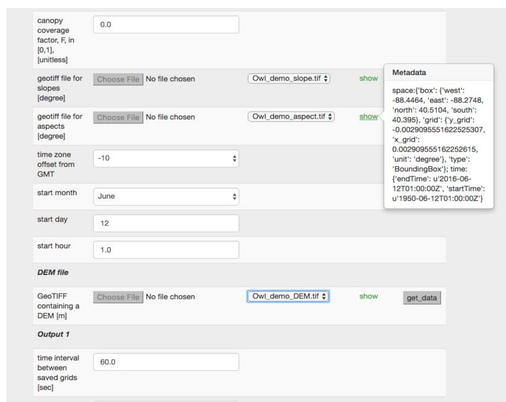
- A platform for integrating BMI-enabled web service models
- It relies on:
  - KIS for the semantic search and mediation(e. g. Clowder)
  - RAS for semantic alignment between the models and data information profiles.



3- Semantic Knowledge Discovery

A sample response of calling KIS based on a model graph to search data in Clowder

4- Semantic Resources Alignment



Project Contributions

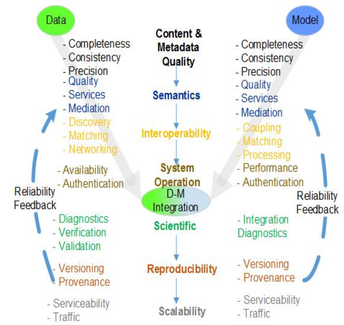
1. Semantically enabled models as a foundation for advancing Model-as-a-Service.
2. EMELI-Web: Web-based model integration engine based on Experimental Modeling Environment for Linking and Interoperability.
3. Graph knowledge base for managing standards and Standard Names.
4. Information system with a SKOS API to create and manage the semantic crosswalks among Standard Names.
5. Semantic Annotation Services for semantic enrichment of data and models.
6. Knowledge Integration Services for ingesting standards and reasoning over their definitions.
7. Resources Alignment Services for handling the mediation between the information profile associated with two geo-resources

Future Path

Going forward: address reliability and consistency challenges in a scalable environment

- Reliability: ability of a scientific workflow to execute correctly and produce scientifically expected results.

- Consistency: ability of these workflows to do so in a non-contradictory manner across instantiations in multiple scientific contexts.



For more details:

Documentation: http://hcgs.ncsa.illinois.edu/index.html

EarthCube Web page: http://earthcube.org/group/geosemantics

Source Code: https://opensource.ncsa.illinois.edu/stash/projects/ECGS