A GeoSemantic Framework for Integrating Long-Tail Data and Models

Luigi Marini, NCSA

Mostafa Elag, UIUC

Praveen Kumar, UIUC
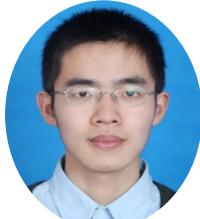
Scott Peckham, CU

Leslie Hsu, LDEO

Rui Lui, NCSA

Peishi Jiang, UIUC

Yao Hu, UIUC

Kim Miller, Wyoming
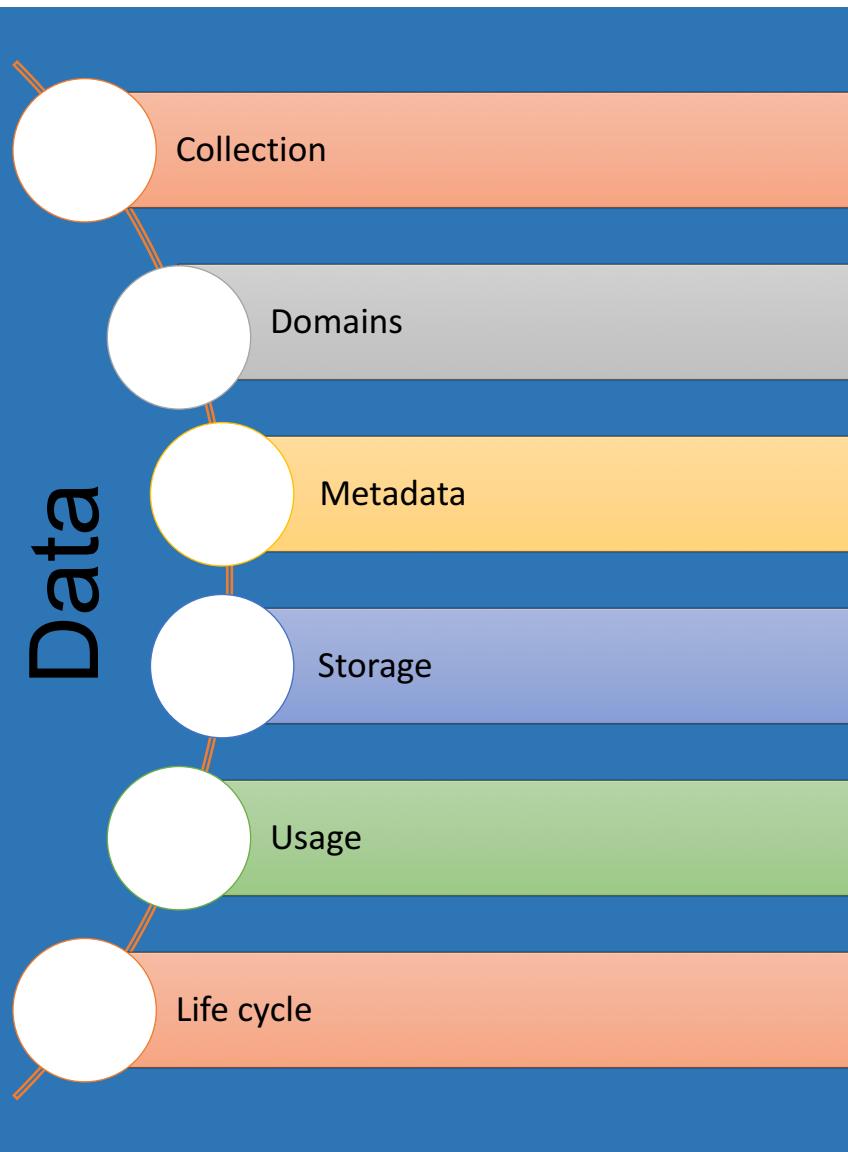
Maria, CU

(Long-Tail) Data-Model Interoperability Challenge

Data

- Collection
- Domains
- Metadata
- Storage
- Usage
- Life cycle

Interoperability Levels

- L0: None
- L1: Technical
- L2: Syntactic
- L3: Semantic
- L4: Pragmatic
- L5: Dynamics
- L6: Conceptual
- L7: Automatic
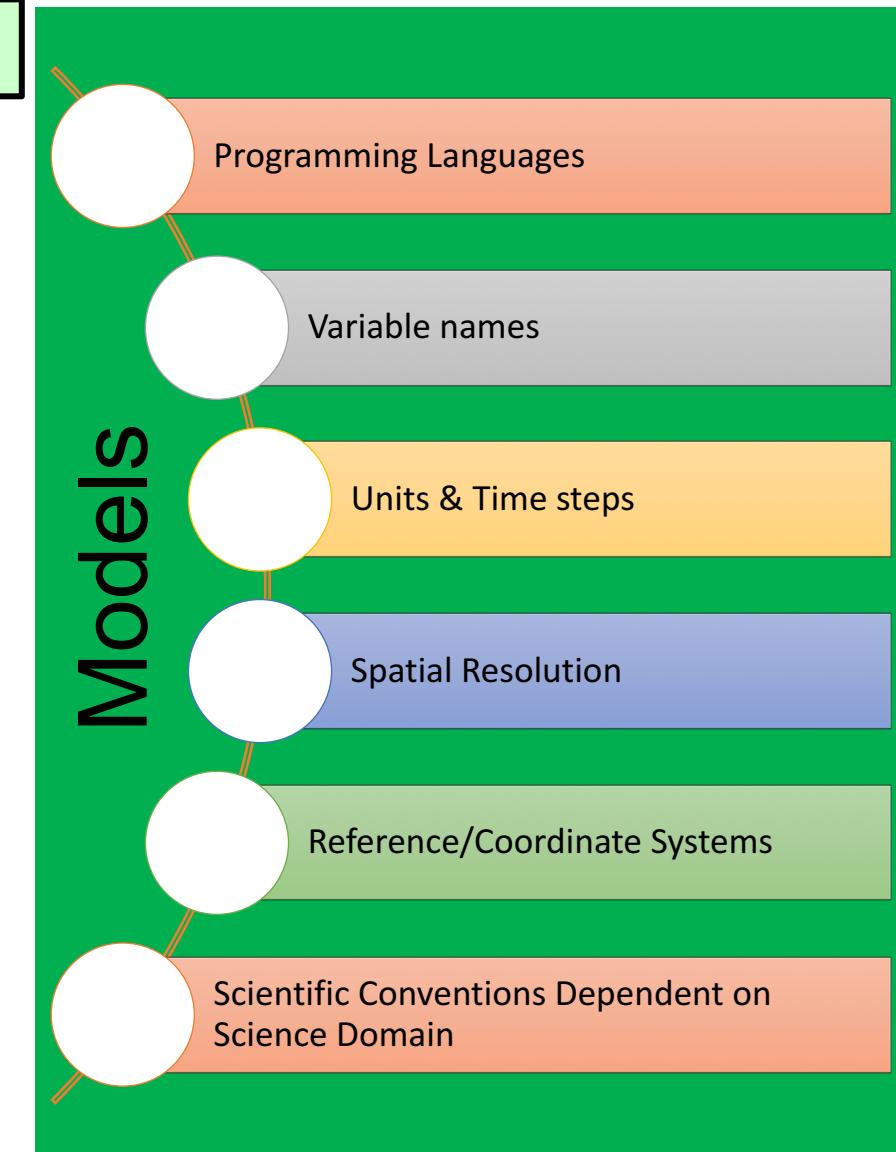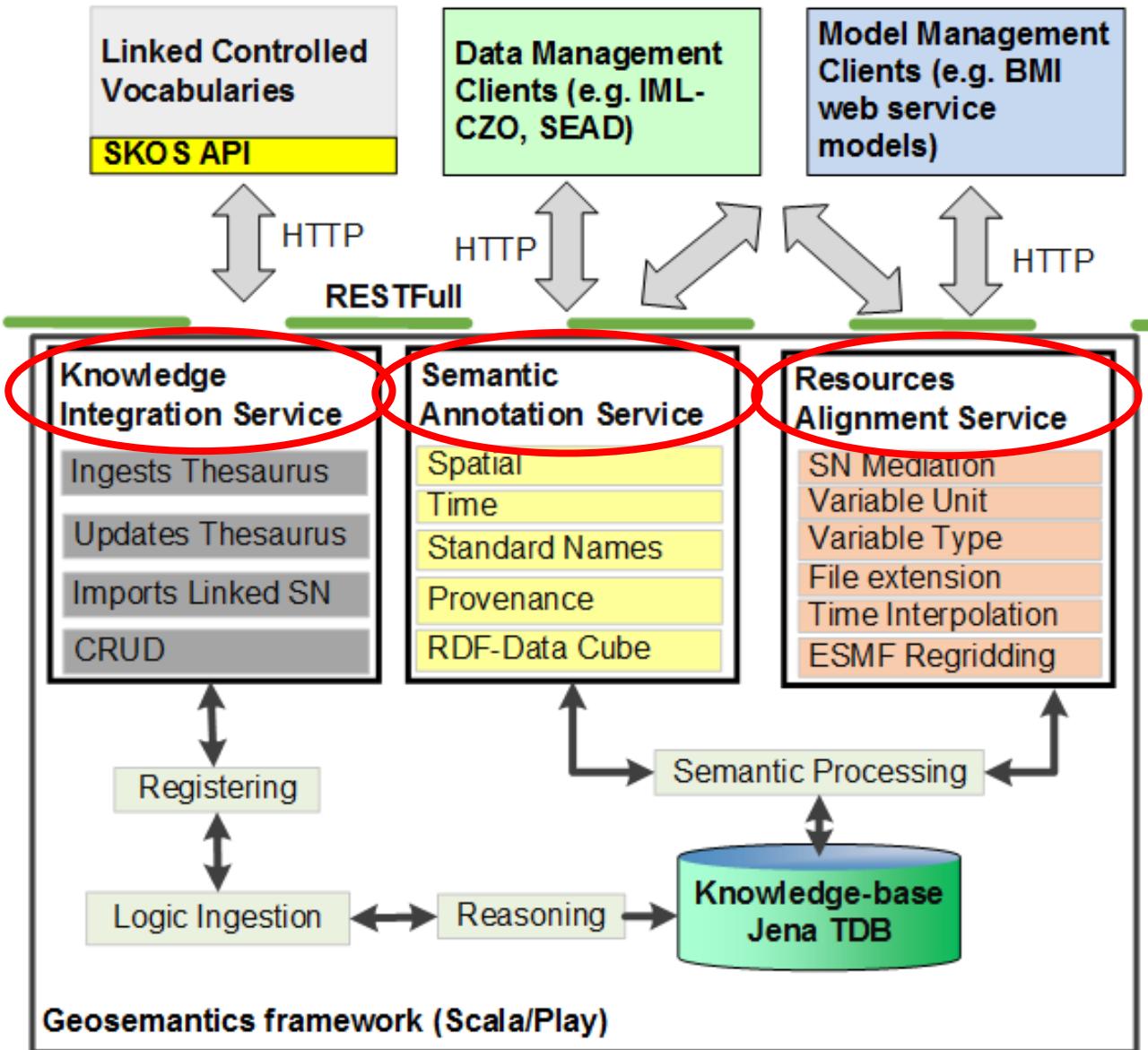
adapted from Wang, et al., 2009

Models

- Programming Languages
- Variable names
- Units & Time steps
- Spatial Resolution
- Reference/Coordinate Systems
- Scientific Conventions Dependent on Science Domain
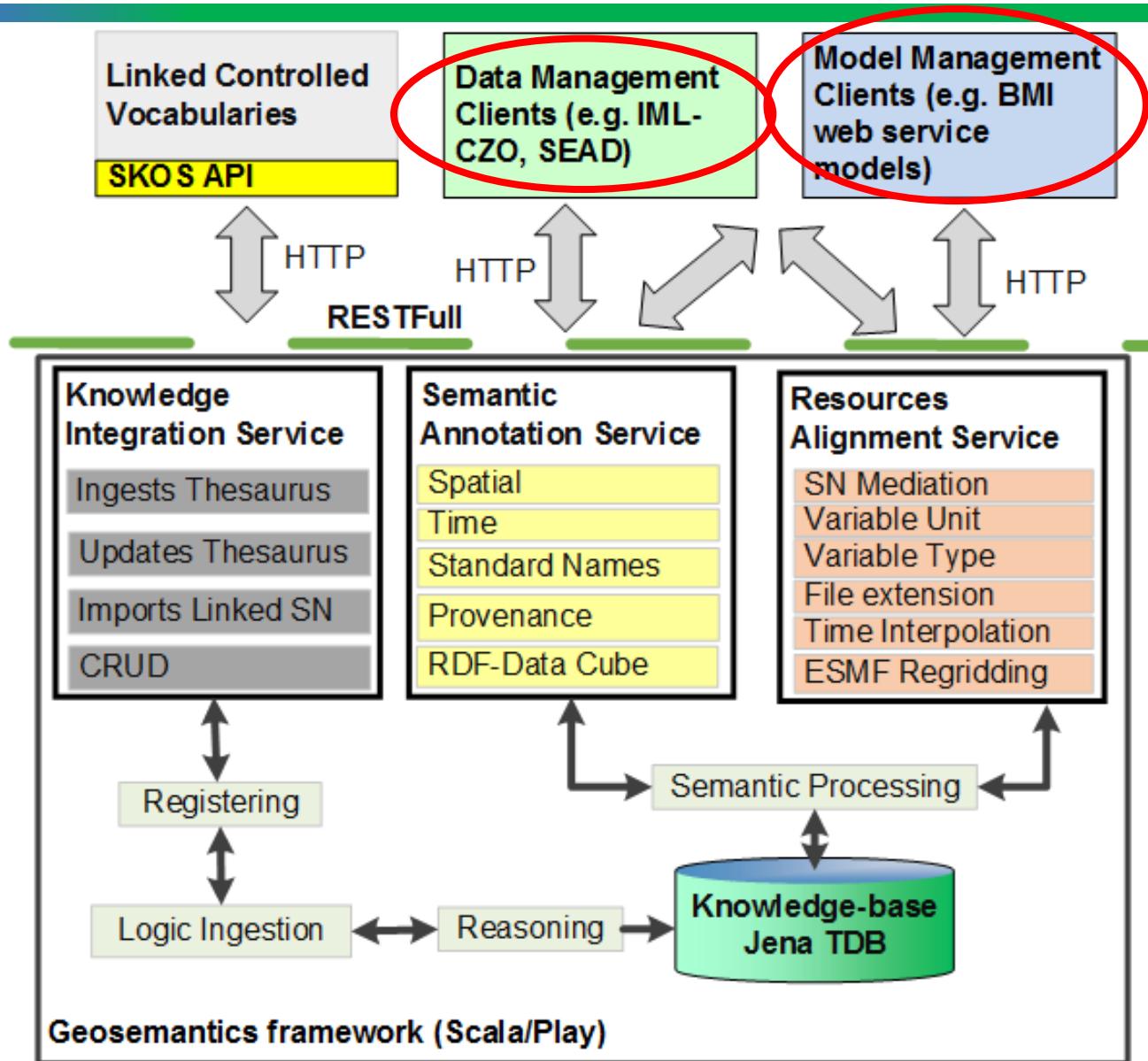
GeoSemantic Approach: OPEN API

- GeoSemantics framework uses Micro-service architecture and Linked Data standards.

- Semantic Annotation Service (SAS)
  1. Annotates resources with spatiotemporal context, variable name, and provenance relationships
  2. Automatic extractors based on the data files MIME type (e.g. GeoTIFF and CSV types)
  3. Provides an interactive interface for manual annotation

- Knowledge Integration Service (KIS)
  1. Ingests, registers, and checks-in Controlled Vocabularies and W3C standards to the framework's Knowledge-base
  2. Provides semantic federated search

- Resource Alignment Service (RAS)
  1. Aligns the information profile associated with two geo-resources to ensure their semantic consistency before integration

GeoSemantic Approach: OPEN API

- Model-as-a-service
  - Allows for rapid integration of heterogeneous models
  - Initial development using BMI enabled CSDMS models
  - EMELI-Web: Web based model integration engine based on Experimental Modeling Environment for Linking and Interoperability - EMELI (Peckham, 2014)
- Leverage and contribute to related technologies/projects
  - SEAD
  - CLOWDER
  - BrownDog
  - IMLCZO
- Demonstration
  - IMLCZO (Intensively Managed Landscape Critical Zone Observatory) context
 - SEN (Sediment Experimentalists Network)
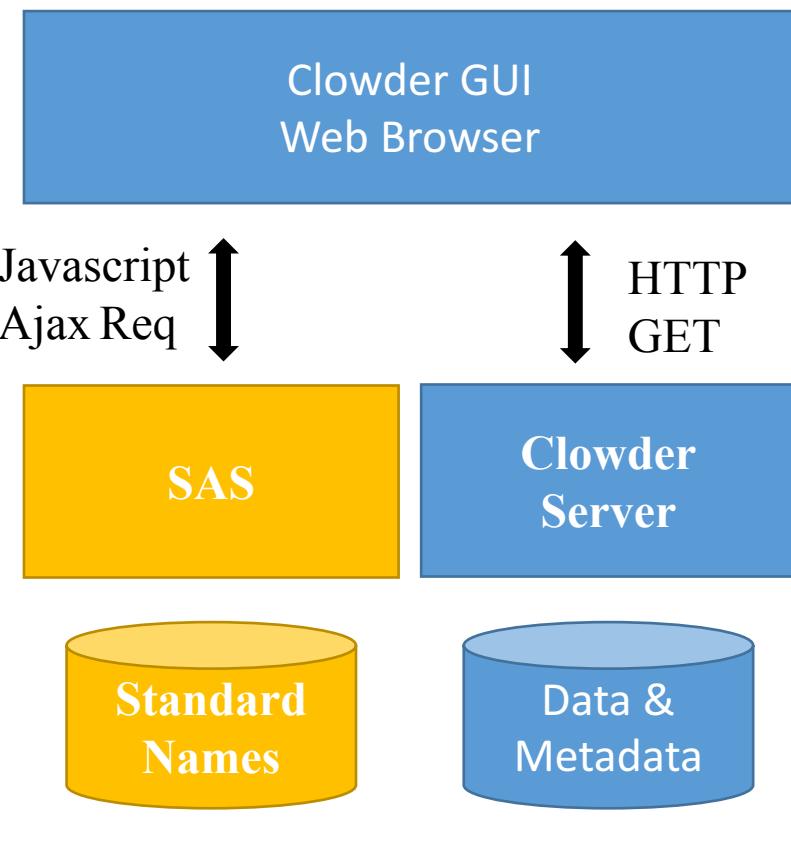
Clowder: Data Management Client

Open source data management for research: https://clowder.ncsa.illinois.edu/

➢ User annotates files and datasets using local definitions and external standard vocabularies defined in SAS

➢ Clowder is one example of a data management system calling SAS. Other data management system can do the same.

Summary and Future Path

➢ GeoSemantic framework provides the services required for seamless semantic integration between data and models.

    ➢ Enables model integration with distributed heterogeneous data resources

    ➢ Enables library of models interoperable

    ➢ Enables data discovery and synthesis, and data analytics

➢ Going forward: address reliability and consistency challenges in a scalable environment using GeoSciences Semantic Infrastructure

    ➢ 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.

THANK YOU!

Resources

- GeoSemantic Technology Descriptions (SAS, RAS, KIS, EMELI-Web, …):

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

    - http://earthcube.org/group/geosemantics