



# Envisioning a Future of Computational Geoscience in a Data Rich Semantic World



Praveen Kumar<sup>a</sup>, Mostafa M. Elag<sup>a</sup>, Peishi Jiang<sup>a</sup>, Luigi Marini<sup>a</sup>, Scott D. Peckham<sup>b</sup>

Leslie Hsu<sup>c</sup>, and Kim Miller<sup>c</sup>

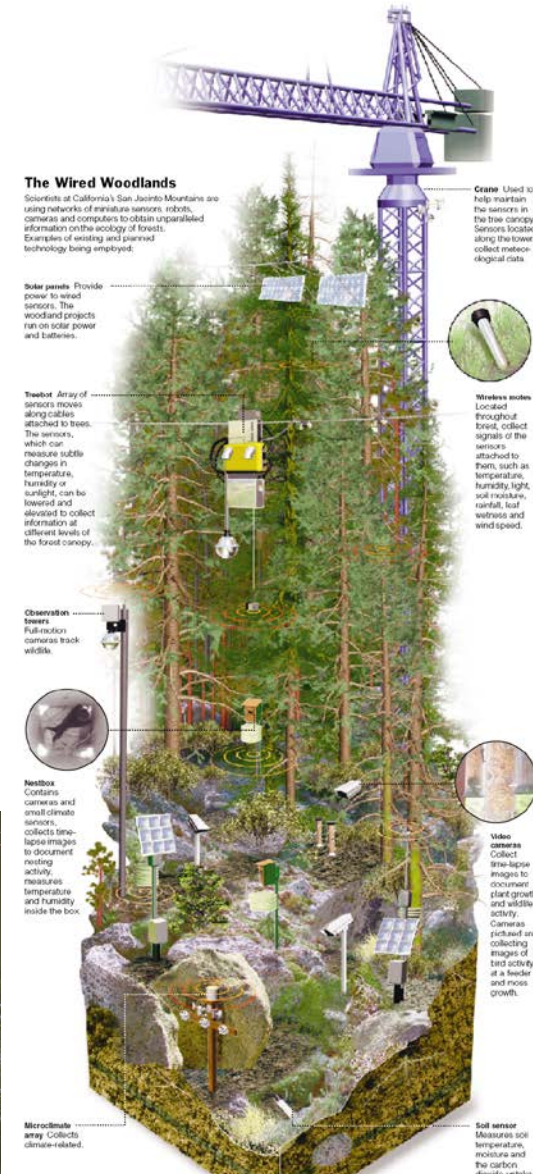
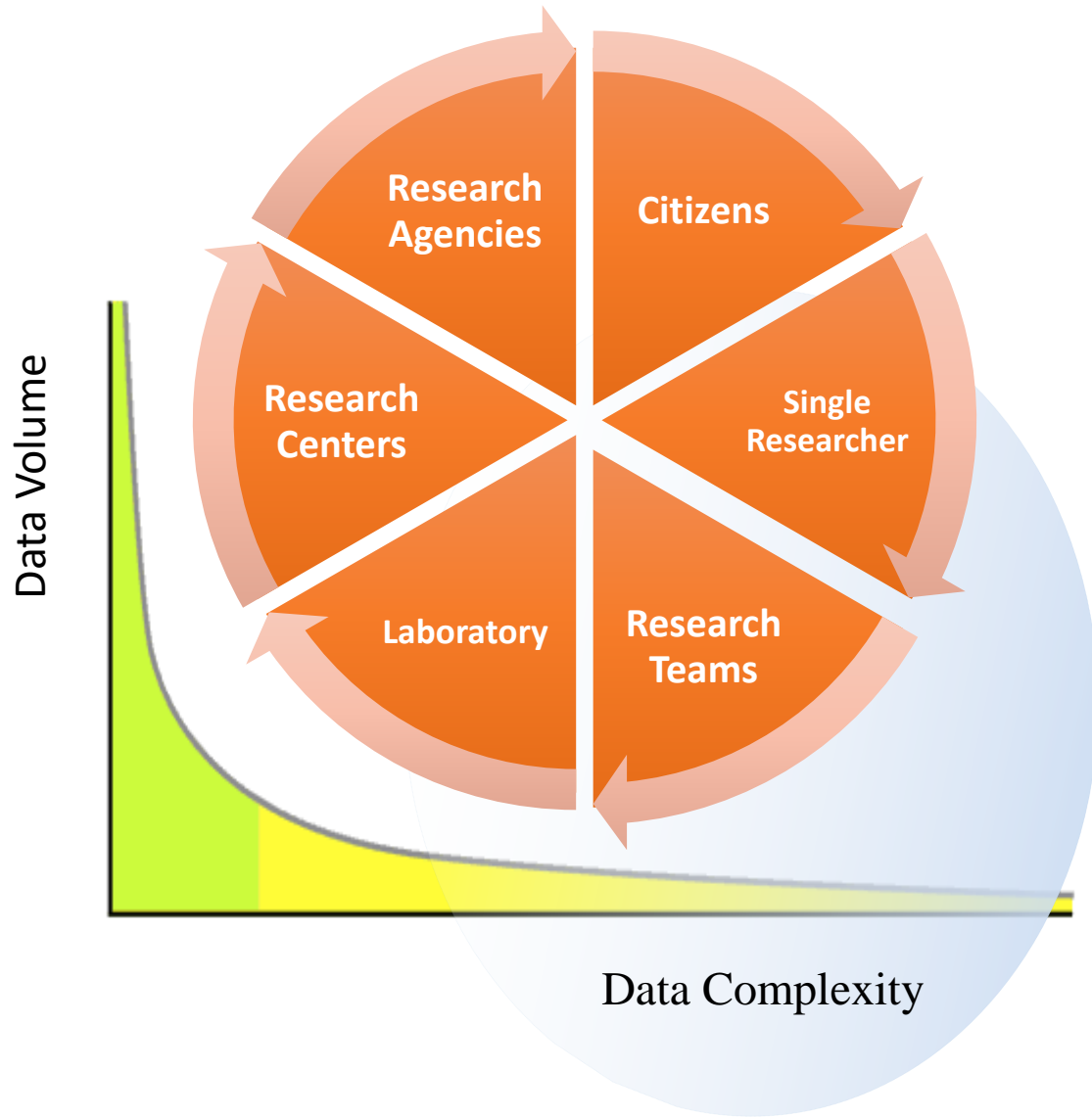
<sup>a</sup>University of Illinois at Urbana-Champaign

<sup>b</sup>University of Colorado Boulder

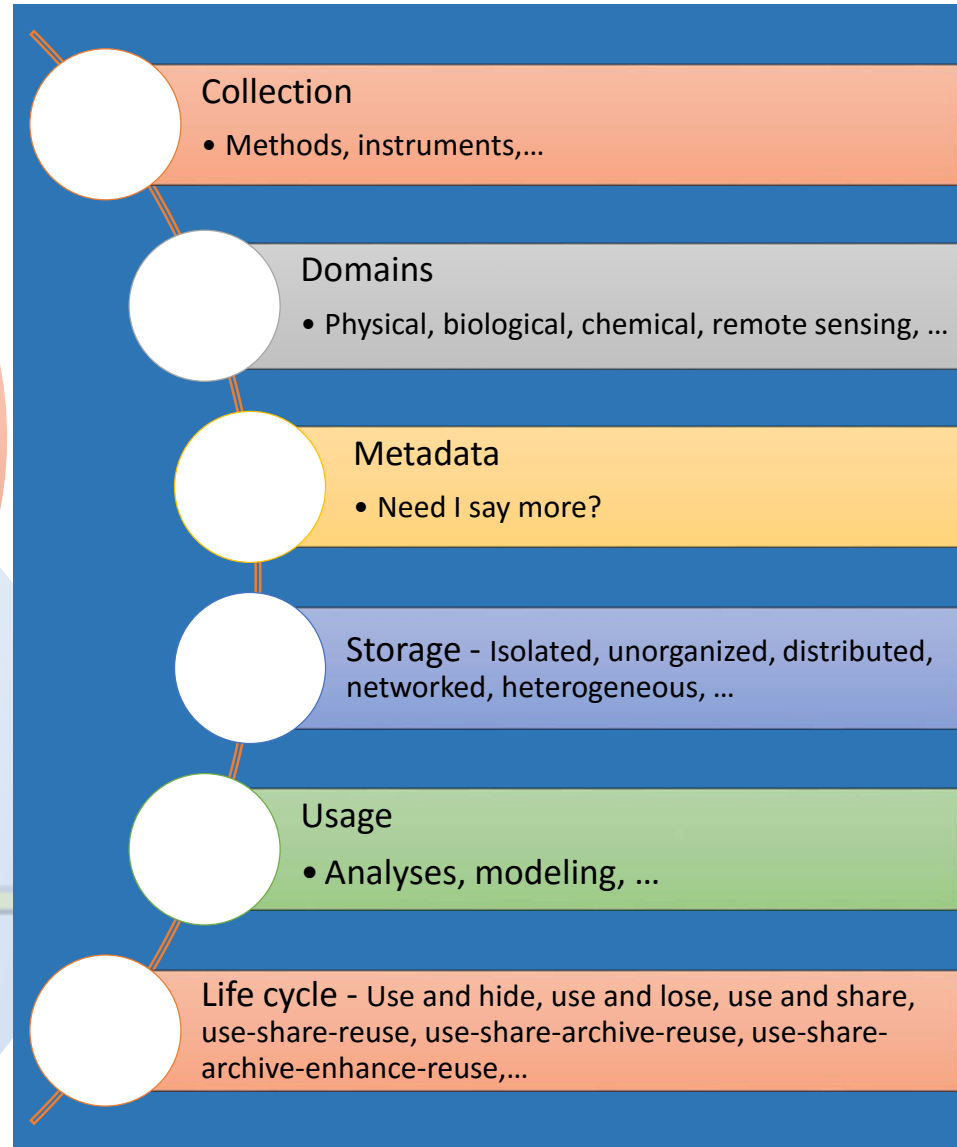
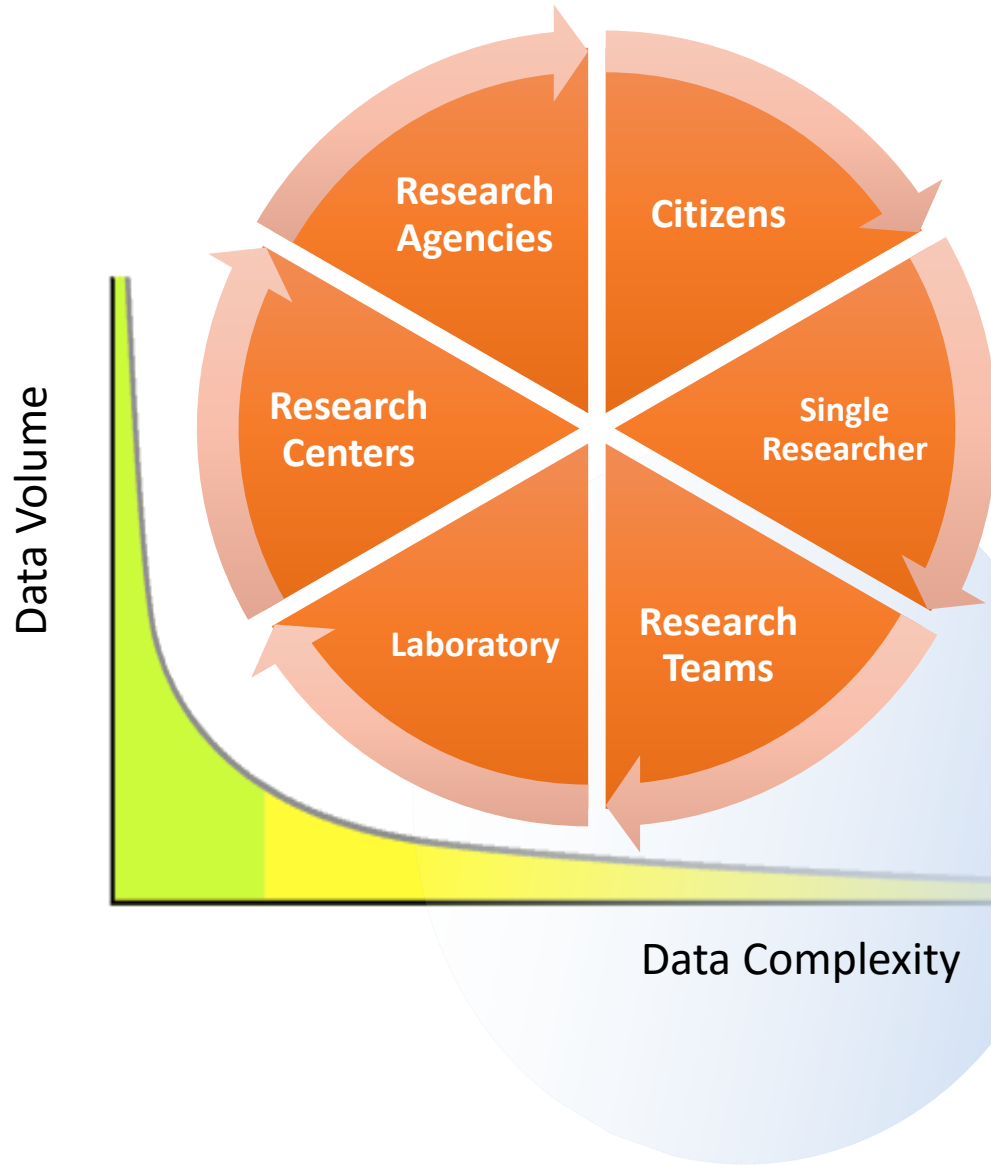
<sup>c</sup>Columbia University



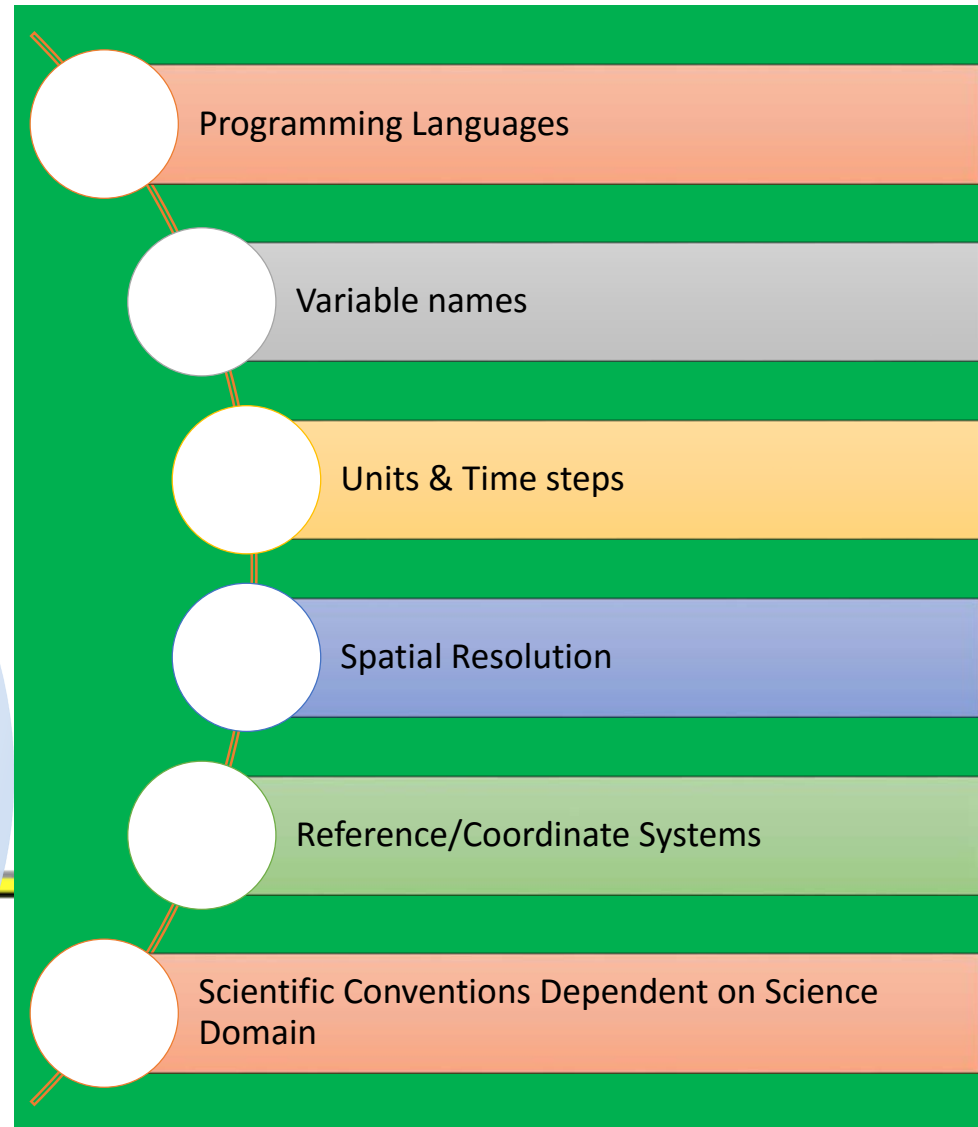
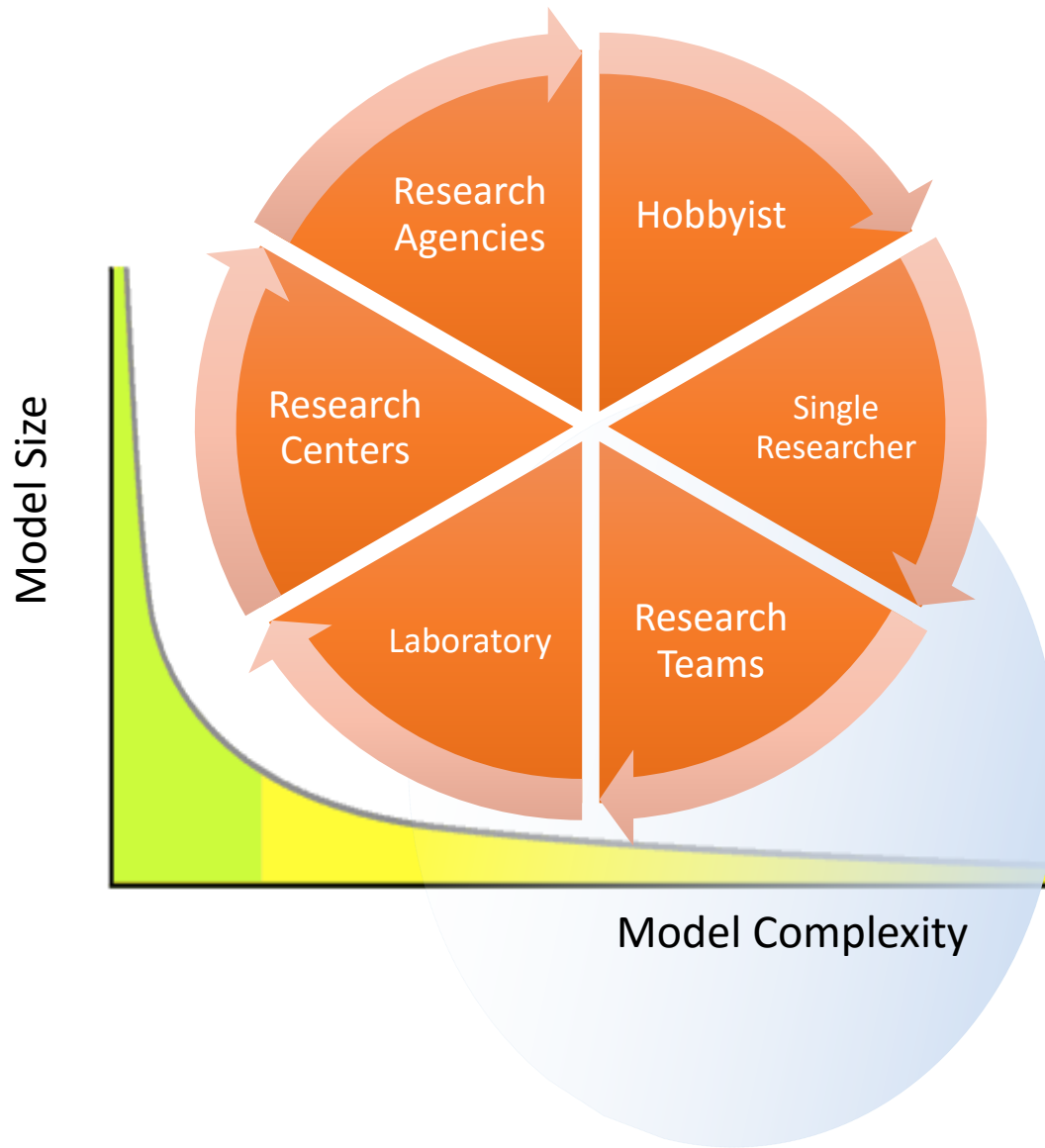
# Complexity of Data: Synthesis and Reuse



# Complexity of Data: Synthesis and Reuse



# Complexity of Models: Interoperability and Coupling



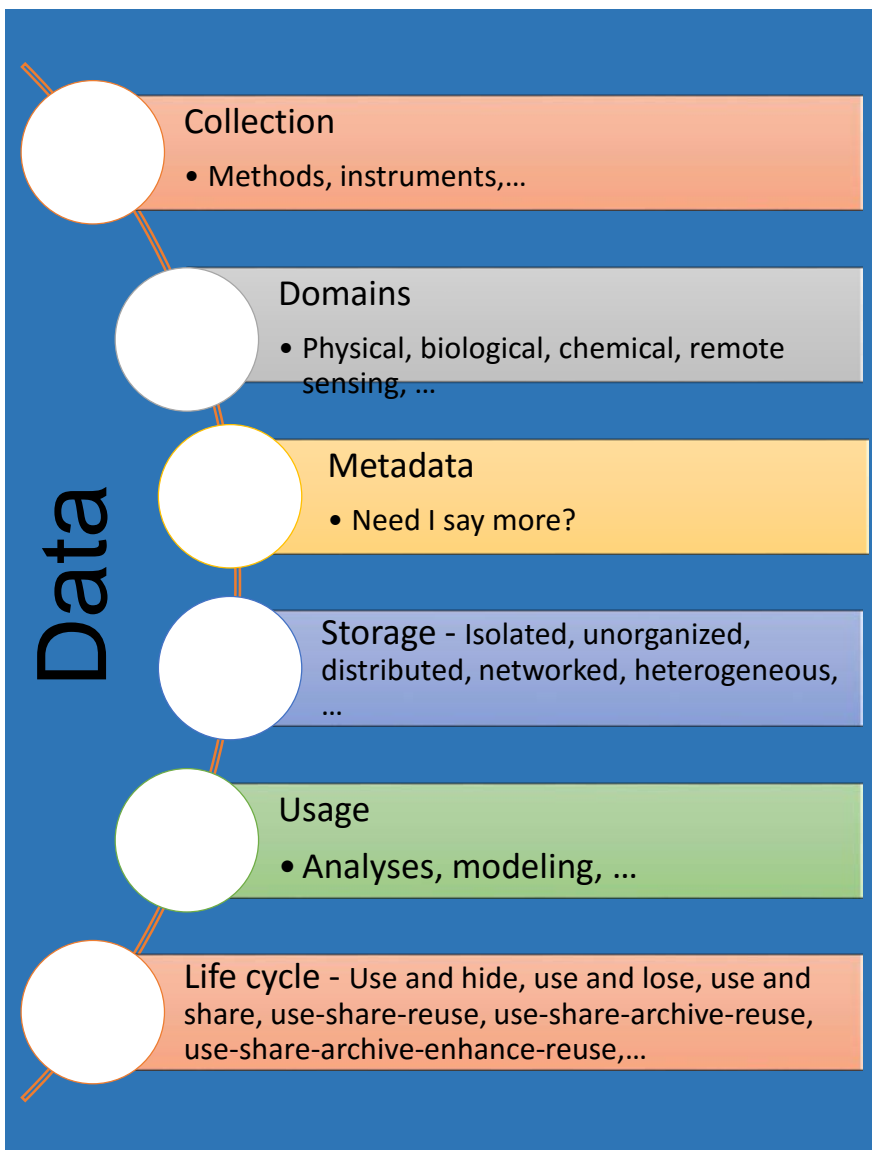
Modeling frameworks

Standalone Models

Model-as-a-Service (MaaS)

Modules and Libraries

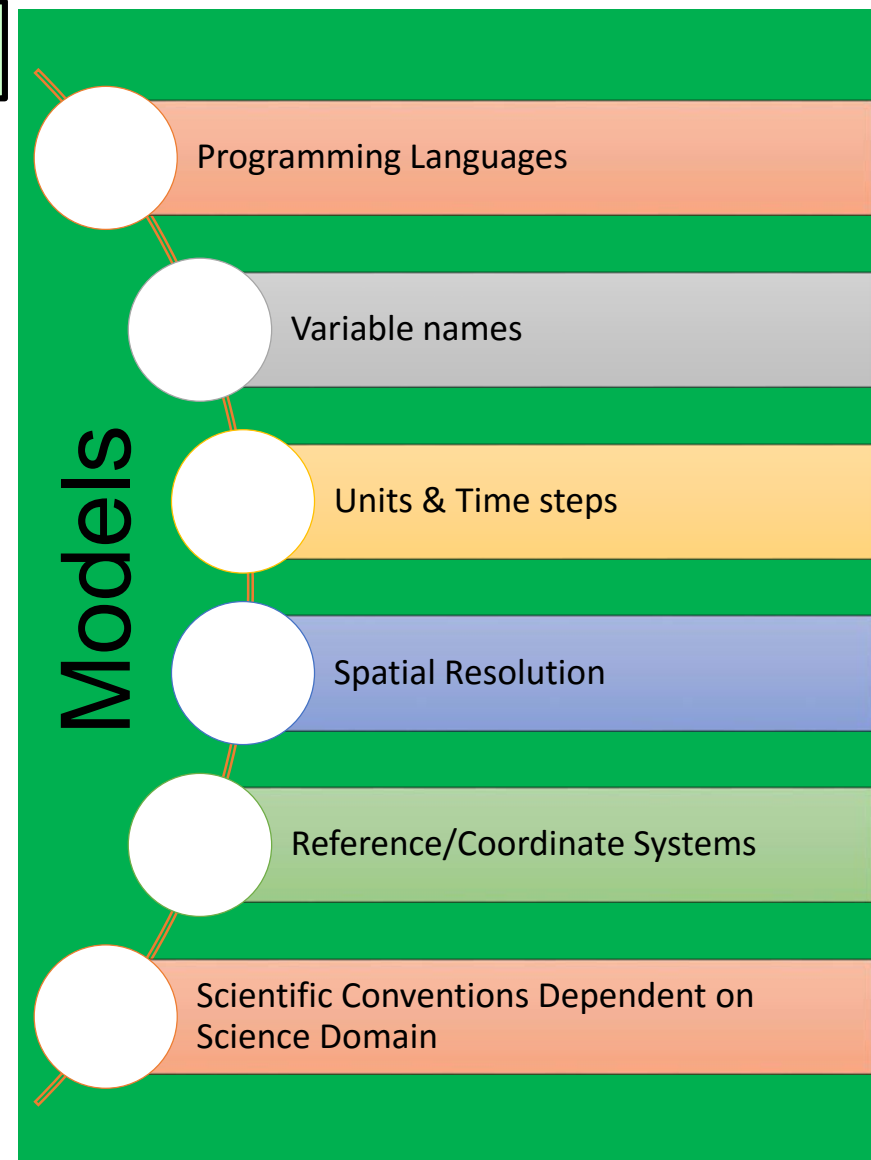
# Data-Model Interoperability



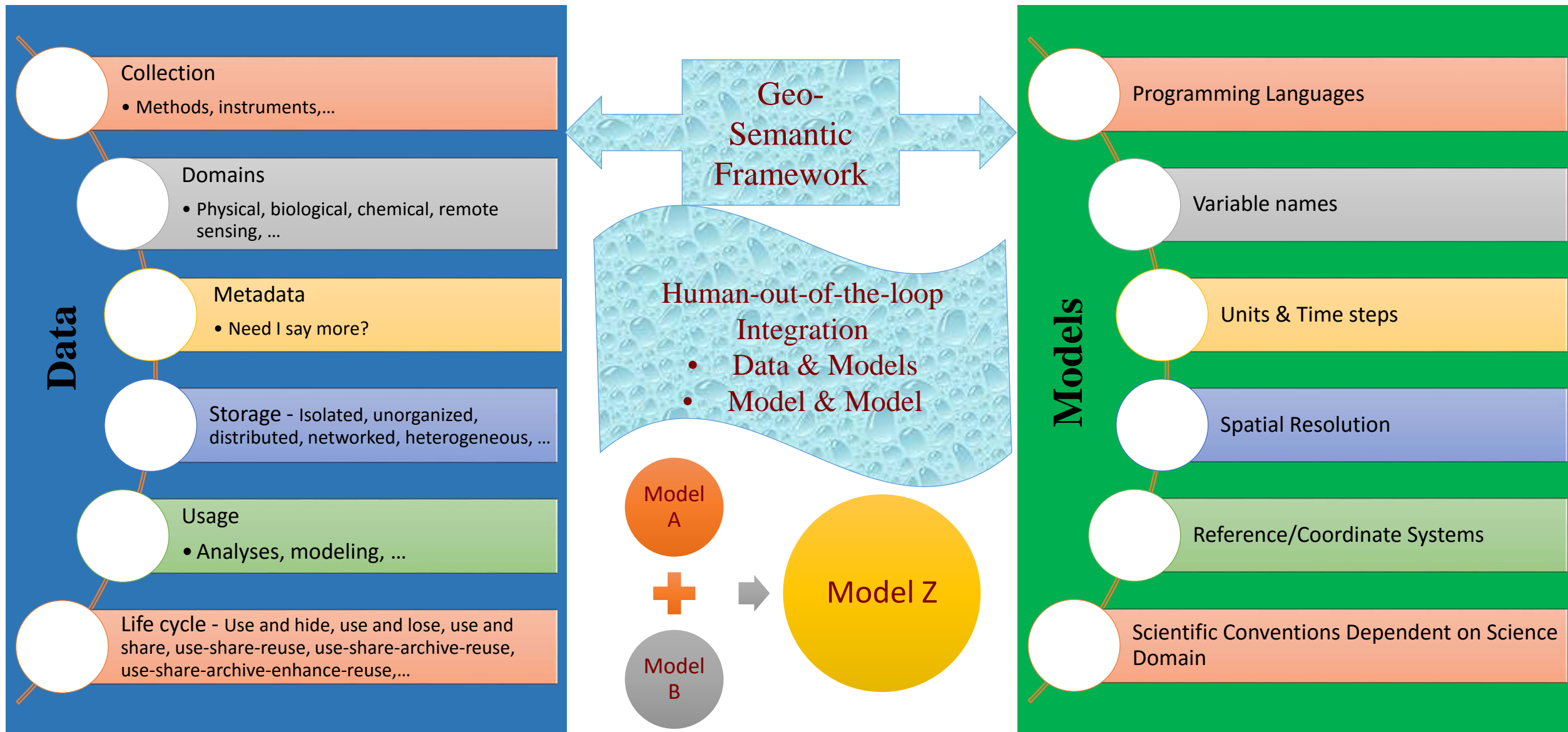
## Interoperability Levels



adapted from Wang, et al., 2009



# Data-Model Semantic Integration



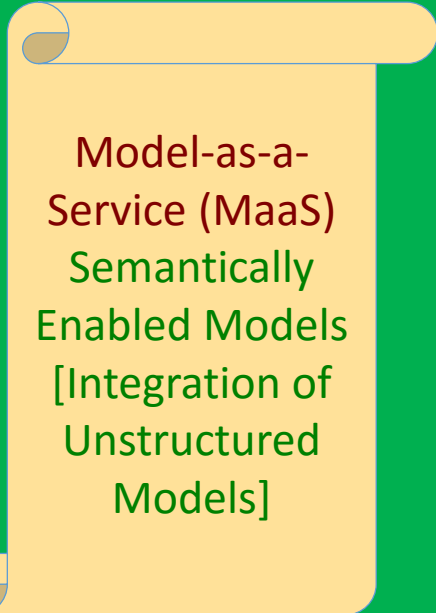

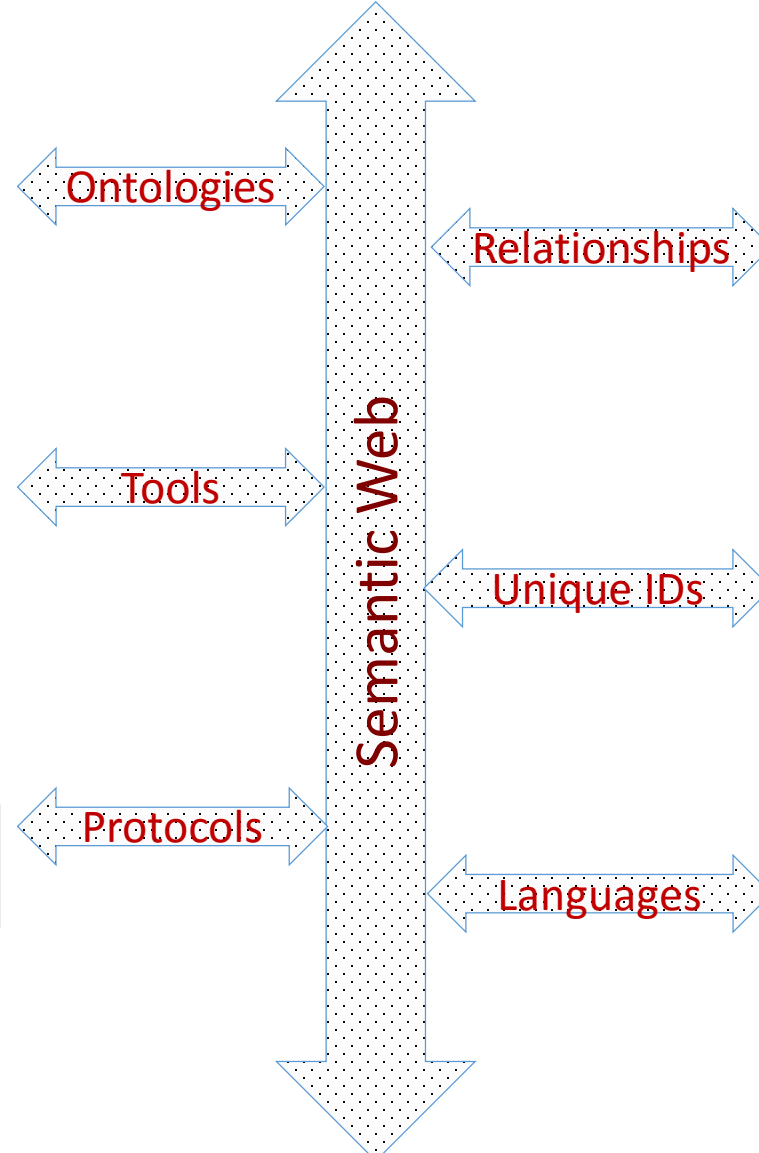
# Data-Model Semantic Integration



**Current**



**OTHERS**



**Current**

**OTHERS**

# Data-Model Semantic Integration

SEAD | Sustainable Environment Actionable Data

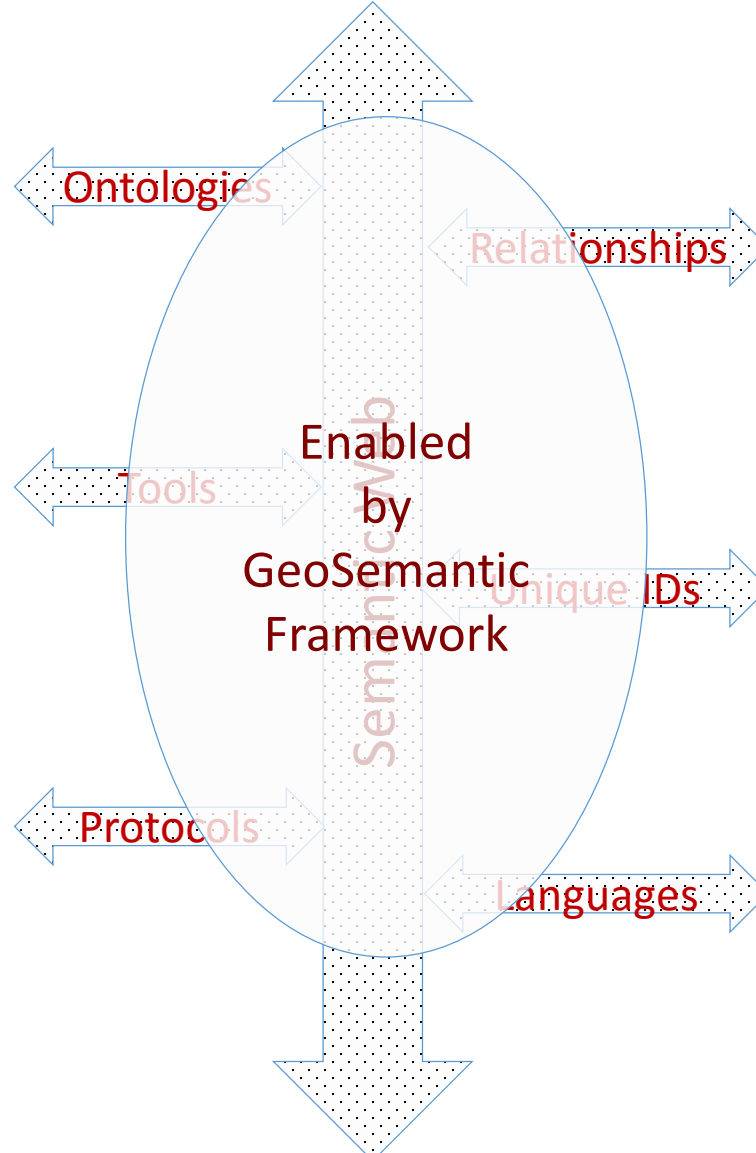
Clowder

**Current**

CUAHSI  
UNIVERSITIES ALLIED FOR WATER RESEARCH

HYDROSHARE

**OTHERS**



CSDMS  
COMMUNITY SURFACE DYNAMICS MODELING SYSTEM

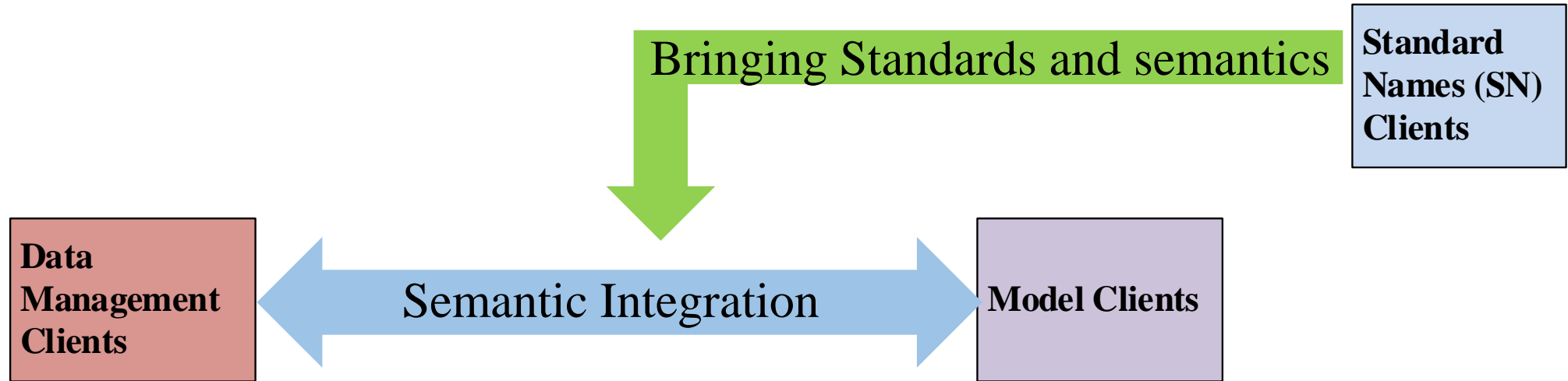
Model-as-a-Service (MaaS)  
Semantically Enabled Models  
[Integration of Unstructured Models]

**Current**

**OTHERS**



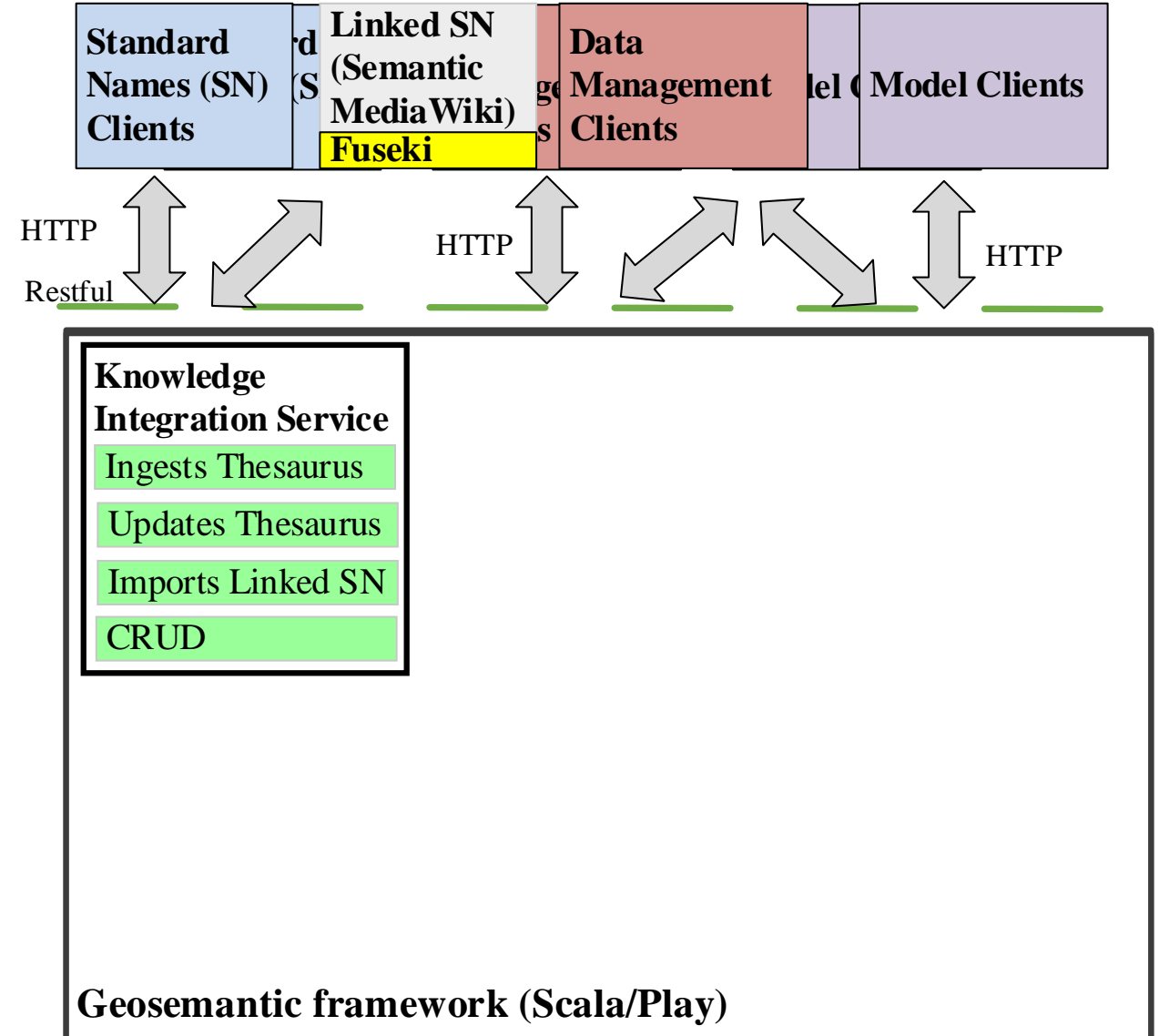
# GeoSemantic Framework Architecture



GeoSemantic framework bridges the gap between Semantic Web standards and resources life cycle by lowering the technology barrier for incorporating semantic in their life cycle.

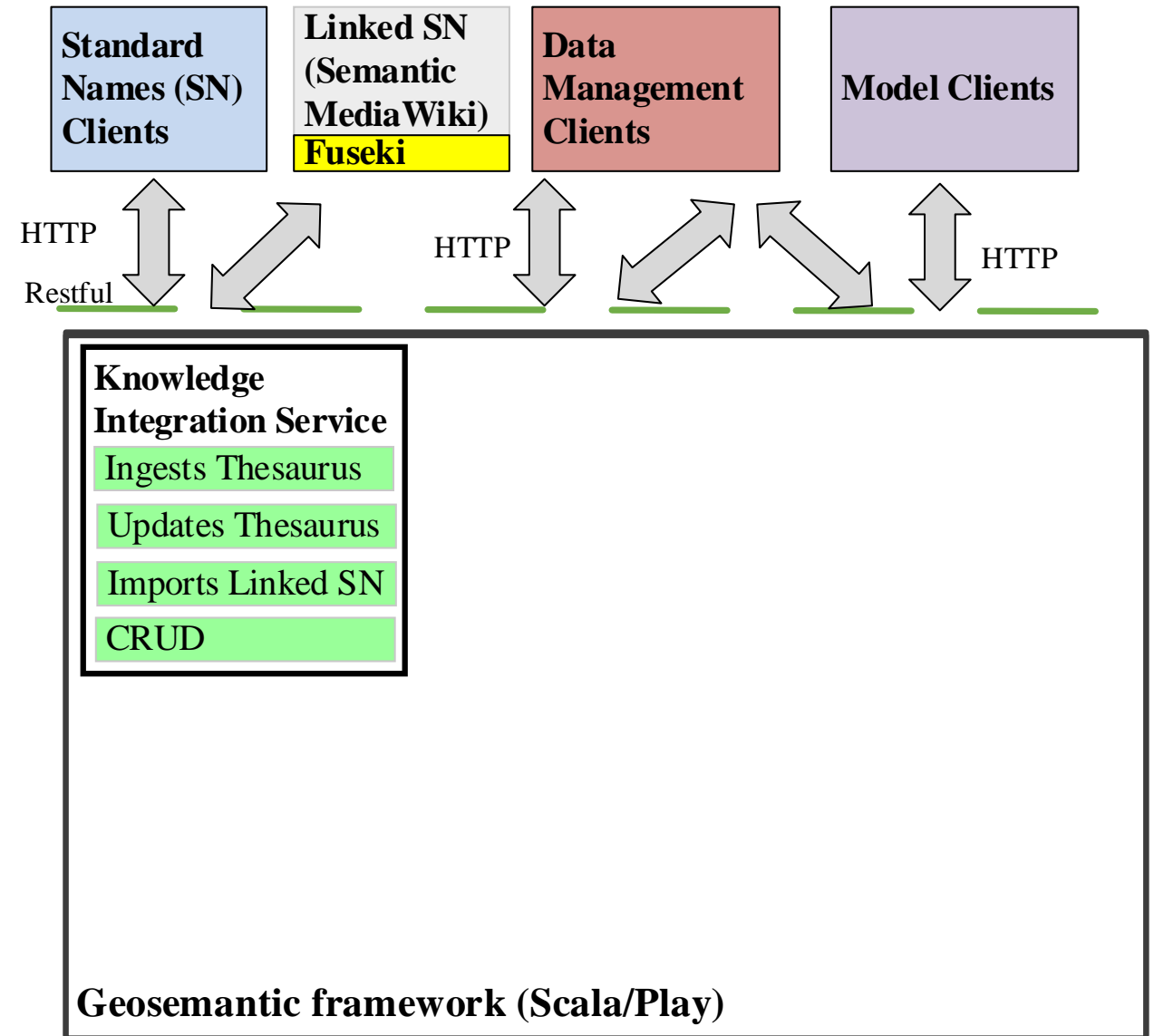
# GeoSemantic Framework Architecture

- It uses Micro-service architecture, where each service focuses on specific function and has a unique endpoint.
- Knowledge Integration Services: It is the information entry point for the framework. It ingests standards and can annotate them back.
- Semantic MediaWiki is used to create semantic relationships among SN.



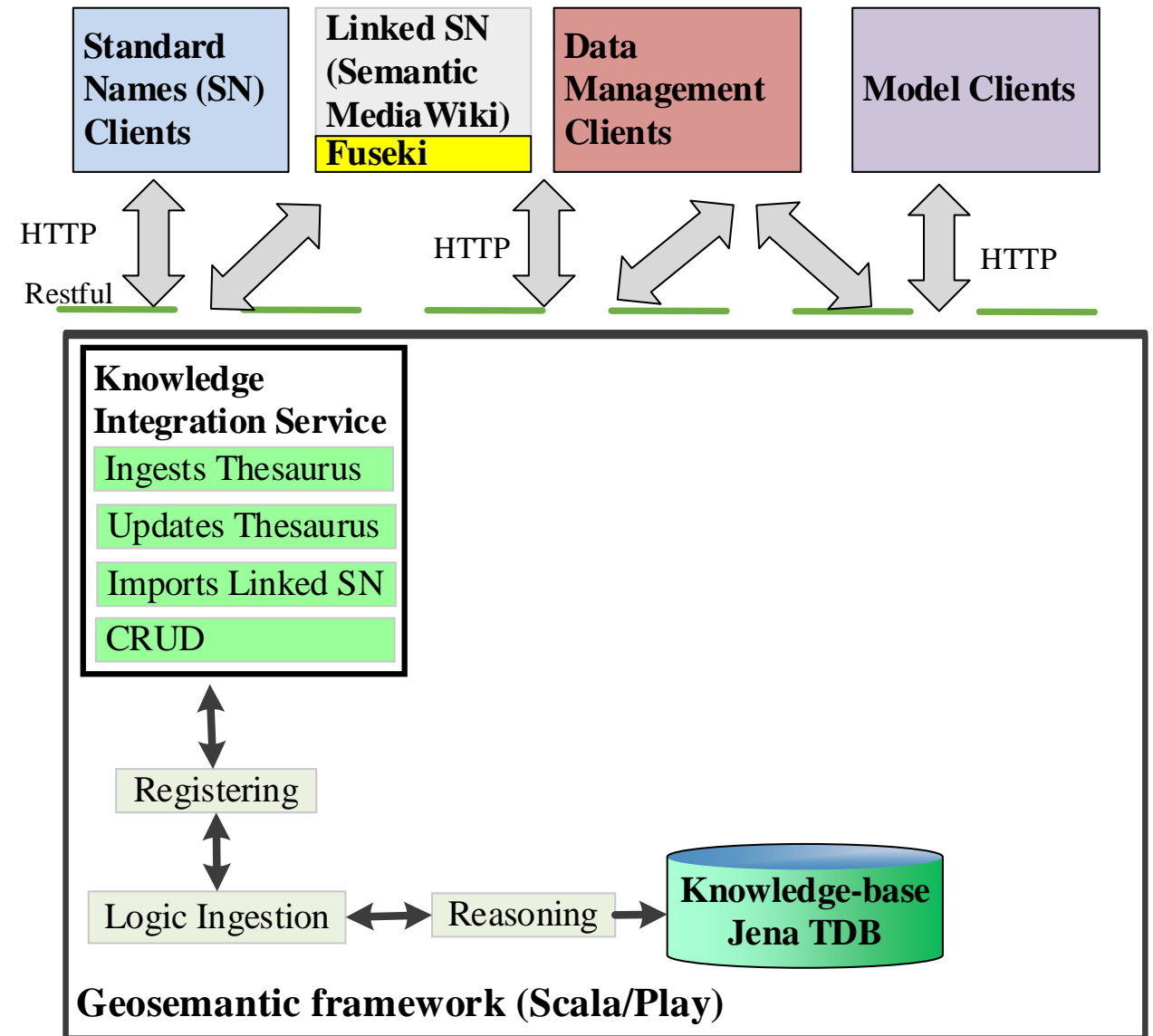
# GeoSemantic Framework Architecture

- It uses Micro-service architecture, where each service focus on specific function and has a unique endpoint.
- Knowledge Integration Services: It is the information entry point for the framework. It ingests standards and can annotate them back.
- Semantic MediaWiki is used to create semantic relationships among SN.



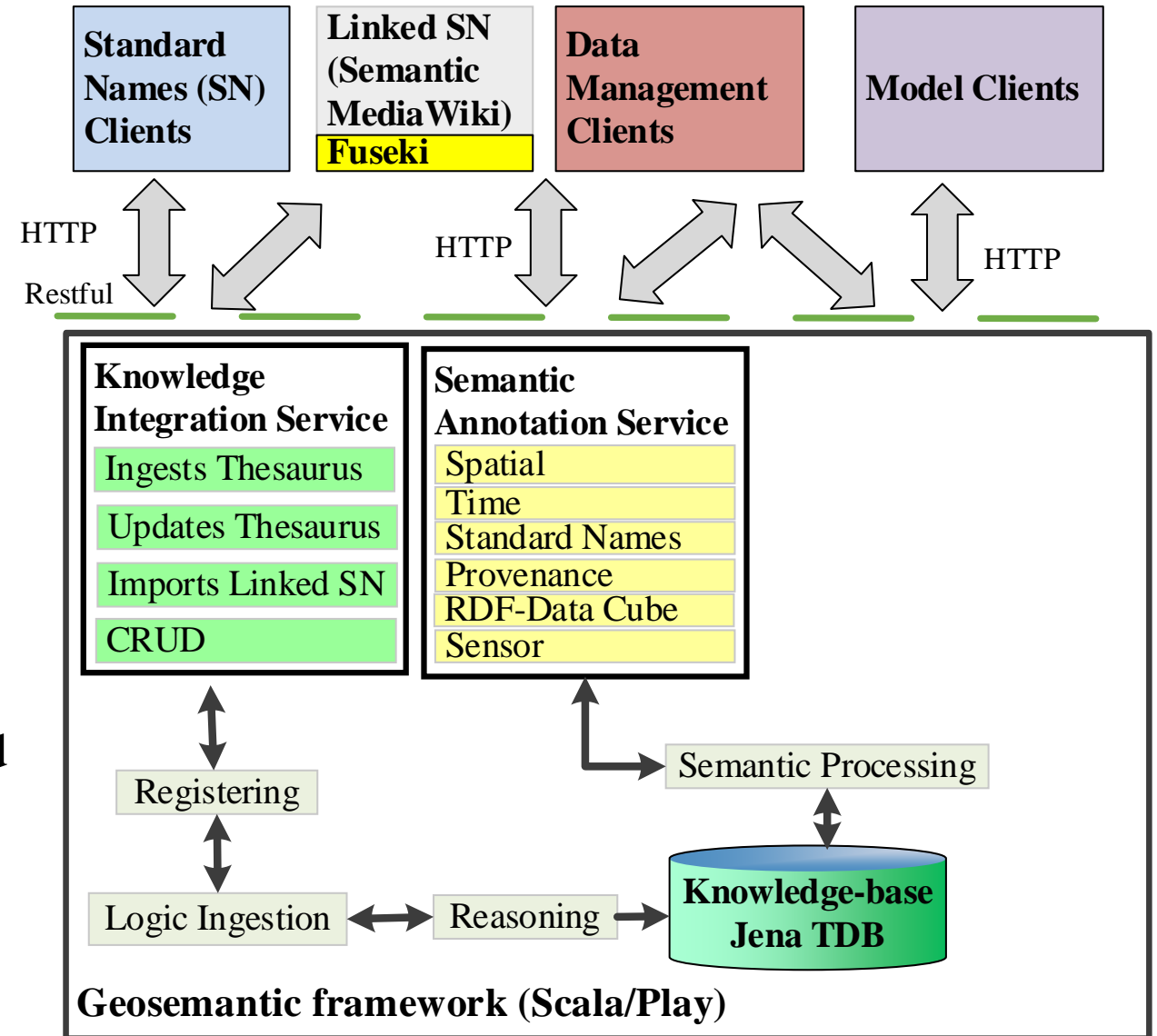
# GeoSemantic Framework Architecture

- It uses Micro-service architecture, where each service focus on specific function and has a unique endpoint.
- Knowledge Integration Services: It is the information entry point for the framework. It ingests standards and can annotate them back.
- Semantic MediaWiki is used to create semantic relationships among SN.
- knowledge base: stores the standard graphs after processing.



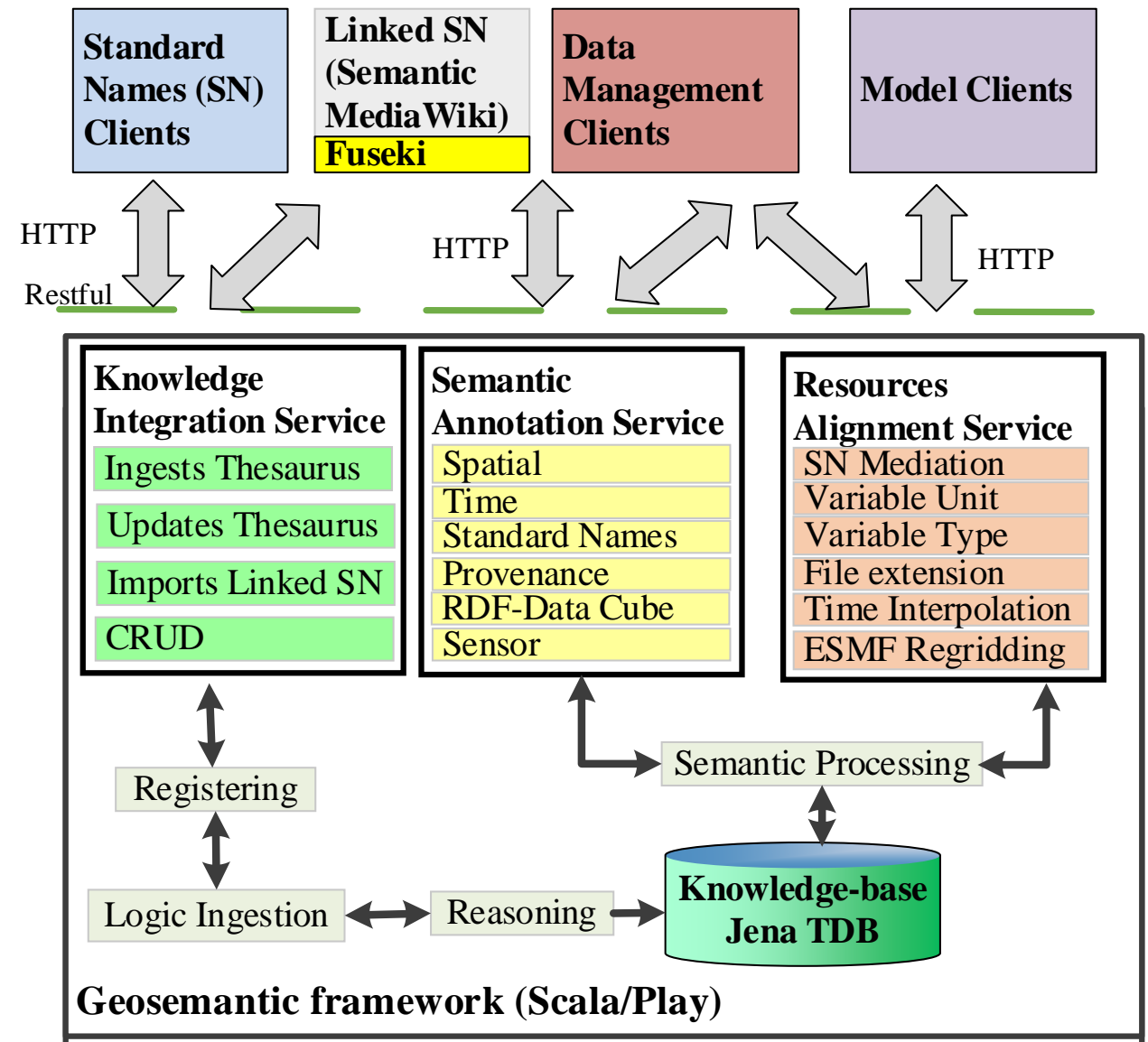
# GeoSemantic Framework Architecture

- It uses Micro-service architecture, where each service focus on specific function and has a unique endpoint.
- Knowledge Integration Services: It is the information entry point for the framework. It ingests standards and can annotate them back.
- Semantic MediaWiki is used to create semantic relationships among SN.
- knowledge base: stores the standard graphs after processing.
- Semantic Annotation Services: enrich the data and models using scientific annotations.



# GeoSemantic Framework Architecture

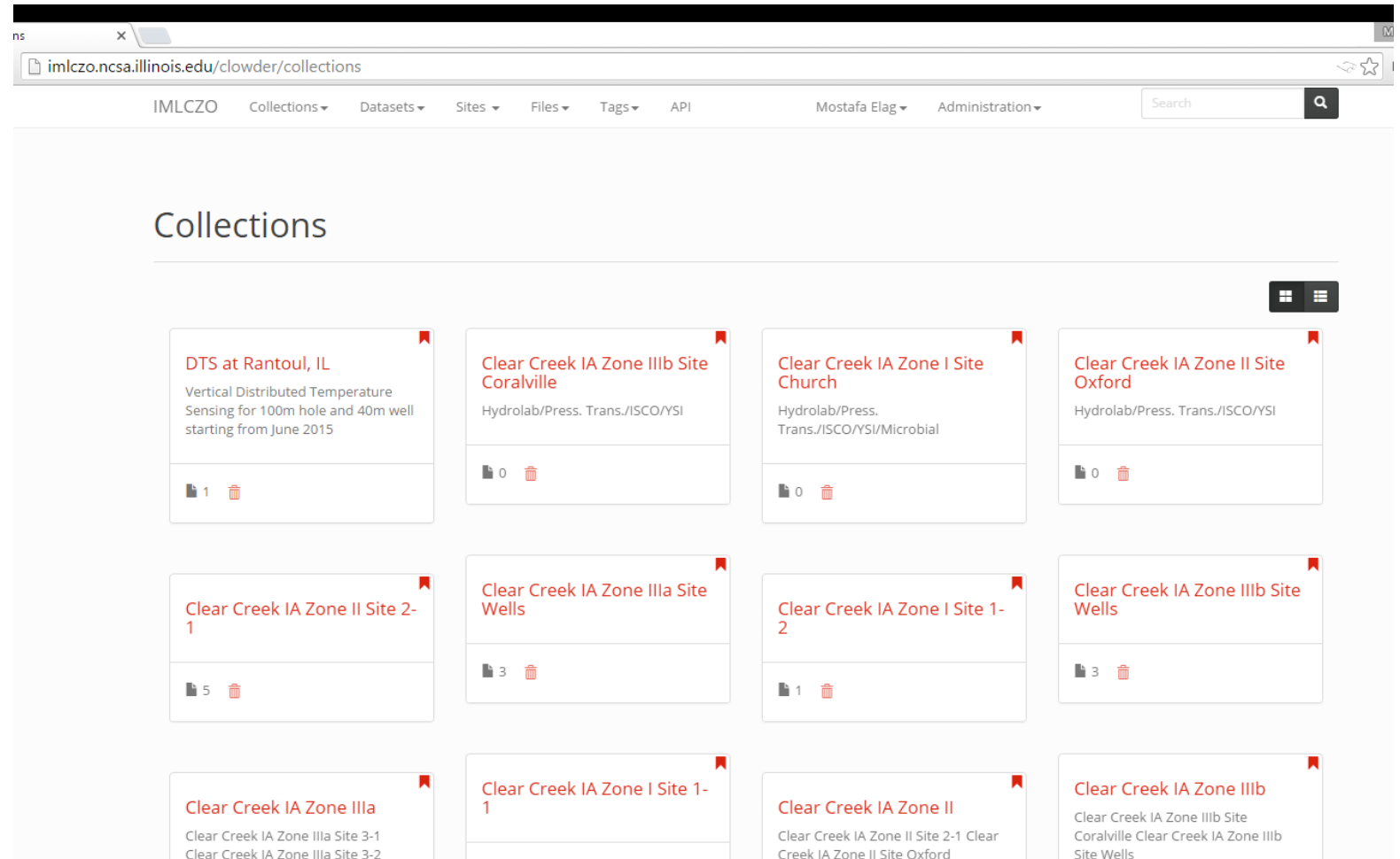
- It uses Micro-service architecture, where each service focus on specific function and has a unique endpoint.
- Knowledge Integration Services: It is the information entry point for the framework. It ingests standards and can annotate them back.
- Semantic MediaWiki is used to create semantic relationships among SN.
- knowledge base: stores the standard graphs after processing.
- Semantic Annotation Services: enrich the data and models using scientific annotations.
- Resource Alignment Services: ensures semantic consistency between coupled models.



# Clients: IML-CZO data and Serviced BMI models

## Semantic enrichment of IML-CZO data

- SEAD is a client for the framework.



The screenshot displays a web interface for the IMLCZO data portal. The browser address bar shows the URL `imlczo.ncsa.illinois.edu/clowder/collections`. The navigation menu includes links for Collections, Datasets, Sites, Files, Tags, and API. The main content area is titled "Collections" and features a grid of 12 collection cards. Each card contains the following information:

- Collection Name:** DTS at Rantoul, IL; Clear Creek IA Zone IIIb Site Coralville; Clear Creek IA Zone I Site Church; Clear Creek IA Zone II Site Oxford; Clear Creek IA Zone II Site 2-1; Clear Creek IA Zone IIIa Site Wells; Clear Creek IA Zone I Site 1-2; Clear Creek IA Zone IIIb Site Wells; Clear Creek IA Zone IIIa; Clear Creek IA Zone I Site 1-1; Clear Creek IA Zone II; Clear Creek IA Zone IIIb.
- Description:** Vertical Distributed Temperature Sensing for 100m hole and 40m well starting from June 2015; Hydrolab/Press. Trans./ISCO/YSI; Hydrolab/Press. Trans./ISCO/YSI/Microbial; Hydrolab/Press. Trans./ISCO/YSI; Hydrolab/Press. Trans./ISCO/YSI; Hydrolab/Press. Trans./ISCO/YSI; Hydrolab/Press. Trans./ISCO/YSI/Microbial; Hydrolab/Press. Trans./ISCO/YSI; Clear Creek IA Zone IIIa Site 3-1; Clear Creek IA Zone IIIa Site 3-2; Clear Creek IA Zone II Site 2-1; Clear Creek IA Zone II Site Oxford; Clear Creek IA Zone IIIb Site Coralville; Clear Creek IA Zone IIIb Site Wells.
- File Count:** 1, 0, 0, 0, 5, 3, 1, 3, 0, 0, 0, 3.

# Clients: IML-CZO data and Serviced BMI models

## Semantic enrichment of IML-CZO data

- SEAD is a client for the framework.
- IML-CZO data space has over than 200 data collections and consumes various Standard Names

The screenshot shows a web interface for a dataset named "Test Dataset 1". At the top, there are three buttons: "Add Files" (with a plus icon), "Delete" (with a trash icon), and "Follow". Below these are five tabs: "Files (1)", "Metadata (6)", "Visualizations", "Comments (0)", and "Notes". The "Metadata (6)" tab is currently selected. In the top right corner of the interface, there is a share icon. The main content area is titled "Add metadata" and features a search bar labeled "Select field". Below the search bar is a list of metadata fields: "Abstract", "Alternative Title", "Audience", "CSDMS Standard Name", "ODM2 Variable Name", "References", "SAS Spatial Geocode", and "SAS Variable Name". The "SAS Variable Name" option is highlighted in blue.



# Clients: IML-CZO data and Serviced BMI models

## Semantic enrichment of IML-CZO data

- SEAD is a client for the framework.
- IML-CZO space has over than 200 data collections and consumes various Standard Names
- Search across stored SN graphs using our search widget.
- Choose from synonyms based on Linked CVs.

Add metadata

SAS Variable Name

SAS Variable Name: temperature dew point

Submit

**CSDMS Standard Name**

- air\_water~vapor\_\_dew\_point\_temperature
- atmosphere\_air\_water~vapor\_\_dew\_point\_temperature
- atmosphere\_bottom\_air\_water~vapor\_\_dew\_point\_temperature
- atmosphere\_water~vapor\_\_dew\_point\_temperature

ODM2 Variable Name

Added by Luigi

context

Add metadata

SAS Variable Name

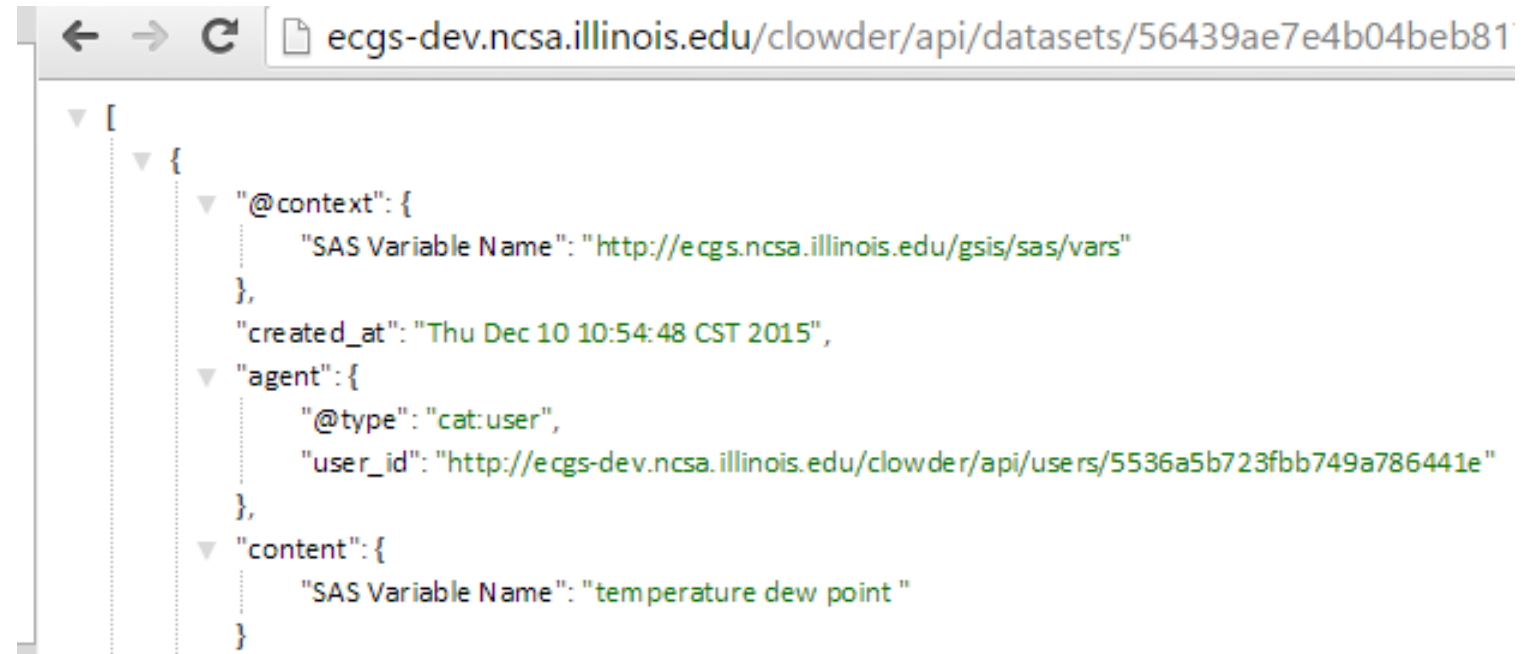
SAS Variable Name: type here...

Use synonyms

# Clients: IML-CZO data and Serviced BMI models

## Semantic enrichment of IML-CZO data

- SEAD is a client for the framework.
- IML-CZO space has over than 200 data collections and consumes various Standard Names
- Search across stored SN graphs using our search widget.
- Choose from synonyms based on Linked CVs.
- All metadata is serialized as JSON-LD



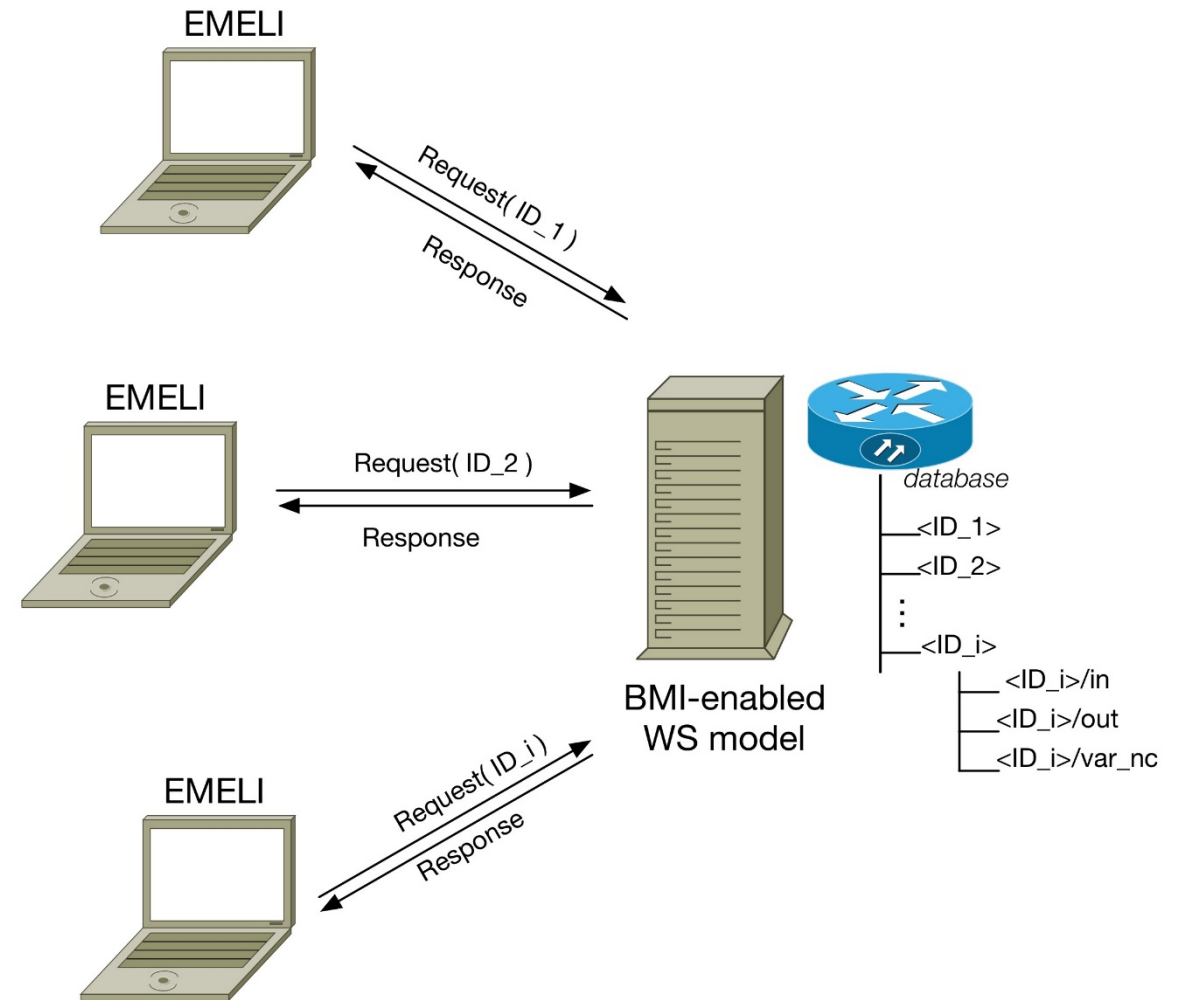
The screenshot shows a web browser window with the address bar containing the URL: `ecgs-dev.ncsa.illinois.edu/clowder/api/datasets/56439ae7e4b04beb81`. The main content area displays a JSON-LD object representing a dataset record. The object is expanded to show its structure:

```
[
  {
    "@context": {
      "SAS Variable Name": "http://ecgs.ncsa.illinois.edu/gsis/sas/vars"
    },
    "created_at": "Thu Dec 10 10:54:48 CST 2015",
    "agent": {
      "@type": "cat:user",
      "user_id": "http://ecgs-dev.ncsa.illinois.edu/clowder/api/users/5536a5b723fbb749a786441e"
    },
    "content": {
      "SAS Variable Name": "temperature dew point "
    }
  }
]
```

# Clients: IML-CZO data and Serviced BMI models

## Serviced BMI models

- BMI-enabled models are exposed as web services
- EMELI 1.0 (Experimental Modeling Environment for Linking and Interoperability) is promoted to EMELI 2.0 to serve Web serviced models.
- A unique UUID is assigned to each model execution task



# Clients: IML-CZO data and Serviced BMI models

## Serviced BMI models

- BMI-enabled models are exposed as web services
- EMELI 1.0 (Experimental Modeling Environment for Linking and Interoperability) is promoted to EMELI 2.0 to serve Web serviced models.
- 
- A unique UUID is assigned to each model execution task.
- Each BMI function has a corresponding route.

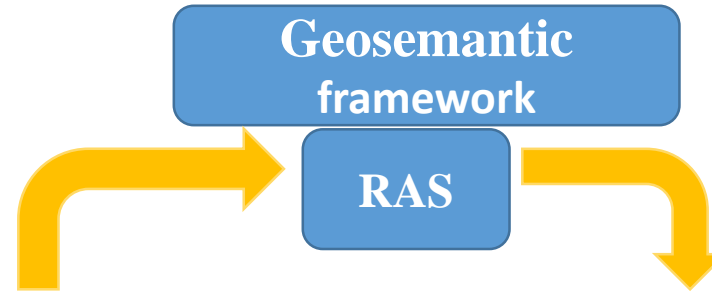
### BMI Functions

BMI Functions	The corresponding routes
	URL <span style="float:right">HTTP method</span>
<b>Model Control Functions</b> + initialize() + update() + finalize()	/models/<model>/instantiate POST /models/<model>/<id>/initialize PUT /models/<model>/<id>/update PUT /models/<model>/<id>/finalize PUT
<b>Model Information Functions</b> + get_input_var_names() + get_output_var_names() + get_attribute()	/models/<model>/<id>/get_input_var_names GET /models/<model>/<id>/get_output_var_names GET /models/<model>/<id>/get_attributes GET
<b>Variable Information Functions</b> + get_var_type()      + get_var_units() + get_var_rank()      + get_var_name() + get_time_step()      + get_time_units() + get_start_time()      + get_current_time() + get_end_time()	/models/<model>/<id>/get_vars_names GET /models/<model>/<id>/get_vars_units GET /models/<model>/<id>/get_time_step GET /models/<model>/<id>/get_time_units GET /models/<model>/<id>/get_time/<when> GET
<b>Variable Getter and Setter Functions</b> + get_value() + set_value()	/models/<model>/<id>/set_values_for_vars PUT
<b>Grid Information Functions</b> + get_grid_spacing() + get_grid_shape() + get_grid_lower_left_corner()	/models/<model>/<id>/<long_var_name>/get_grid_properties GET

# Integration between IML-CZO data and Serviced BMI models

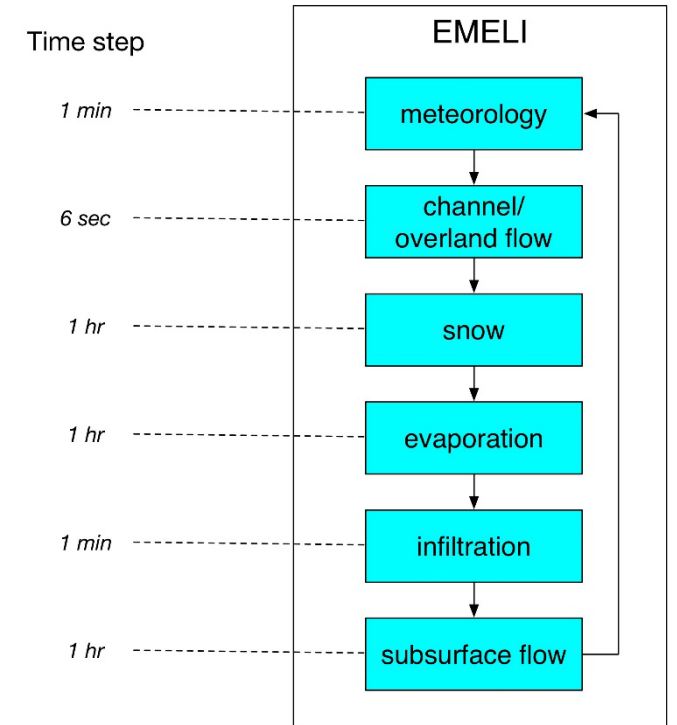
## Goose creek watershed, IML-CZO

```
{
  "created_at": "Tue Nov 24 10:24:22 CST 2015",
  "agent": {
    "@type": "cat:user",
    "user_id": "http://e.cgs-dev.ncsa.illinois.edu/clowder/api/users/555a0c983f22ab256761eab0"
  },
  "content": {
    "SAS Variable Name": "Abs"
  }
},
{
  "@context": {
    "SAS Spatial Geocode": "http://e.cgs.ncsa.illinois.edu/gsis/sas/geocode"
  },
  "created_at": "Wed Nov 18 09:58:24 CST 2015",
  "agent": {
    "@type": "cat:user",
    "user_id": "http://e.cgs-dev.ncsa.illinois.edu/clowder/api/users/553f9fc123fbb749a786443a"
  },
  "content": {
    "Location Name": "Denver, CO, USA",
    "Latitude": "39.74",
    "Longitude": "-104.99"
  }
},
{
  "@context": {
    "SAS Variable Name": "http://e.cgs.ncsa.illinois.edu/gsis/sas/vars"
  },
  "created_at": "Wed Nov 18 09:57:37 CST 2015",
  "agent": {
    "@type": "cat:user",
    "user_id": "http://e.cgs-dev.ncsa.illinois.edu/clowder/api/users/553f9fc123fbb749a786443a"
  },
  "content": {
    "SAS Variable Name": "odm2:windGustSpeed"
  }
},
{
  "@context": {
    "Alternative Title": "http://purl.org/dc/terms/alternative"
  }
}
```



## Resources Alignment between IML-CZO data and BMI models

## Serviced Topoflow components



# Summary and Future work

- GeoSemantic framework lowers the barrier for incorporating semantics in resource life cycle by:
  1. Ingesting standards and ontologies.
  2. Creating Standard Names semantic wiki for linking Controlled Vocabularies.
  3. Providing SAS for semantic enrichment of resources.
  4. Developing RAS to ensure semantic consistency between coupled resources.
  
- Future work will concentrate on:
  1. Adapting more standards and endpoints to satisfy different geoscience communities.
  2. Converting EMELI to be a web service.
  3. Creating MicroData templates to incorporate semantic annotation directly in HTML.
  4. Developing micro-services to annotate simulation models.

# Project URLs

- We invite input and feedback from the Geoscience community at:
  1. EarthCube: <http://workspace.earthcube.org/geo-semantic>
  2. Confluence: <https://opensource.ncsa.illinois.edu/confluence/display/ECGS/GeoSemantic>
- We encourage developers to contribute to the framework source code at <https://opensource.ncsa.illinois.edu/stash/projects/ECGS>
- Geosemantic Wiki of Standard Names is available at <http://ecgs.ncsa.illinois.edu/mediawiki/index.php/Main>

## Acknowledgments

Support from NSF grants ACI-0940824", ACI-1261582", EAR-1331906", and ICER-1440315" are gratefully acknowledged

THANK YOU