Ontology and Taxonomy Development for GEOSS AR-09-01d

UT Japan, Ryosuke Shibasaki, shiba@csis.u-tokyo.ac.jp
ESA, Sergio D’Elia, Sergio.DElia@esa.int
IEEE, SJS Khalsa, sjsk@nsidc.org
Challenge for Data Interoperability

**Syntax Interoperability**
Proposal of Standard Schema and Interface. It is not always enough for diversified Geo-spatial Information.

**Semantic Interoperability**
Semantic Interoperability for geo-spatial data by using data definitions, terminologies, relations, gazetteers (Ontology) etc.

Need to Manage Heterogeneous Data

- Very Large Image Data
- Data Stream from Sensor Networks
- Integrating Observation Data and Model Simulation
- A1: height growth society
  - A2: diversified type
- B1: sustainable development type
  - B2: regional coexistence type
- Land use model
- Socio Economic data schema
- Health data schema
- Agriculture data schema
- Crop Yield data schema
- Hydrology data schema
Sub-task Definition
(as given in the 2009-2011 Work Plan):

• As part of the **Best Practices Registry**, create an **Ontology and Taxonomy** section to get an overview of available ontologies and taxonomies.

• **Compare and analyze** ontologies and taxonomies such as to avoid unnecessary overlaps and conflicts.

• As appropriate, **develop ontologies and taxonomies** stored in the **Best Practices Registry into standards**.

• **Assist in the deployment of a referenceable ontology** for Earth observation to link the **User Requirements Registry with the Components and Services Registry**.

• **Assess how to use the ontology and taxonomy** section of the best practices registry for discovery composition and access in the frame of the **GEOSS architecture**.
Proposed approach

1. Collect outline information on existing ontologies and taxonomies associated with Earth observation, and provide an overview.

2. Compare and analyze the ontologies and taxonomies.
   + Scope, Objectives, Language?
   + Overlaps? Relationships? Maintenance?
   + Models? (format, logical structure etc.)

3. Populate Best Practice Wiki on Ontologies/Taxonomies as referenceable basis for GEOSS.

4. Assess the possibility of “Standard” Ontology/Taxonomies for GEOSS.

5. Assist in the deployment of the referenceable ontologies and taxonomies in Best Practice Wiki for Earth observation in populating the Registries.

6. Assess how to use the ontologies and taxonomies for discovery composition and access in the frame of the GEOSS architecture.
Proposed approach

1. Collect outline information on existing ontologies and taxonomies associated with Earth observation, and provide an overview.

2. Compare and analyze the ontologies and taxonomies.
   + Scope, Objectives, Language?
   + Overlaps? Relationships? Maintenance?
   + Models? (format, logical structure etc.)

3. Populate Best Practice Wiki on Ontologies/Taxonomies as referenceable basis for GEOSS.

4. Assess the possibility of “Standard” Ontology/Taxonomies for GEOSS.

5. Assist in the deployment of the referenceable ontologies and taxonomies in Best Practice Wiki for Earth observation in populating the Registries.

6. Assess how to use the ontologies and taxonomies for discovery composition and access in the frame of the GEOSS architecture.
Overview of existing ontologies/taxonomies

Starting with the list of available ontology

http://babel.cis.u-tokyo.ac.jp/dict/ontology/index.php/SWEET
http://babel.cis.u-tokyo.ac.jp/rdic2-ontology/
Relationships among ontologies/taxonomies
Examples of Gazetteers

Digital Gazetteer (Global)
- ADL
- Geonames
- Wikimapia

Geospatial Ontology
- The Getty Thesaurus of Geographical names
- GEMET
- Wikipedia

Geocoding Services
- Google Geocoding
- YahooGeocoding
- CSIS Address matching service (Japan)
Examples of Regional Gazetteers

Geoscience Australia (Australia)
Natural Resources Canada (Canada)
China Historical GIS (China)
genealogy.net (Germany)
Base map information view service (Japan)
Proposed approach

1. Collect outline information on existing ontologies and taxonomies associated with Earth observation, and provide an overview.

2. Compare and analyze the ontologies and taxonomies.
   + Scope, Objectives, Language?
   + Overlaps? Relationships? Maintenance?
   + Models? (format, logical structure etc.)

3. Populate Best Practice Wiki on Ontologies/Taxonomies as referenceable basis for GEOSS.

4. Assess the possibility of “Standard” Ontology/Taxonomies for GEOSS.

5. Assist in the deployment of the referenceable ontologies and taxonomies in Best Practice Wiki for Earth observation in populating the Registries.

6. Assess how to use the ontologies and taxonomies for discovery composition and access in the frame of the GEOSS architecture.
GEMET (General Multilingual Environmental Thesaurus)
- Groups / Themes, 21 languages

EARTh (Environmental Applications Reference Thesaurus)
- GEMET for Italy (English – Italian version)

Wiktionary (~ Wikipedia)
- Multilingual dictionary with definitions, etymologies,…

Eurovoc Thesaurus
- On EC fields, 21 languages

SWEET (Semantic Web for Earth and Environmental Terminology)
- Earth science data / information

None focused on EO applications (see also InterRisk)
Proposed approach

1. Collect outline information on existing ontologies and taxonomies associated with Earth observation, and provide an overview.

2. Compare and analyze the ontologies and taxonomies.
   - Scope, Objectives, Language?
   - Overlaps? Relationships? Maintenance?
   - Models? (format, logical structure etc.)

3. Populate Best Practice Wiki on Ontologies/Taxonomies as referenceable basis for GEOSS.

4. Assess the possibility of a “Standard” Ontology/Taxonomy for GEOSS.

5. Assist in the deployment of the referenceable ontologies and taxonomies in Best Practice Wiki for Earth observation in populating the Registries.

6. Assess how to use the ontologies and taxonomies for discovery composition and access in the frame of the GEOSS architecture.
Overview of existing ontologies/taxonomies

Starting with the list of available ontology
Individual terms can also be managed with Wiki.

Relationships among terms

http://babel.cis.u-tokyo.ac.jp/vrdic-awci/
KeyGraph Viewer

http://babel.cis.u-tokyo.ac.jp/vrdic-awci/
Proposed approach

1. Collect outline information on existing ontologies and taxonomies associated with Earth observation, and provide an overview.

2. Compare and analyze the ontologies and taxonomies.
   - Scope, Objectives, Language?
   - Overlaps? Relationships? Maintenance?
   - Models? (format, logical structure etc.)

3. Populate Best Practice Wiki on Ontologies/Taxonomies as referenceable basis for GEOSS.

4. **Assess the possibility of a “Standard” Ontology/Taxonomy for GEOSS.**

5. Assist in the deployment of the referenceable ontologies and taxonomies in Best Practice Wiki for Earth observation in populating the Registries.

6. Assess how to use the ontologies and taxonomies for discovery composition and access in the frame of the GEOSS architecture.
Minimum EO Ontology

Application Term

Product

Service

Processor

Product = Packed data or information

Unstable: many changes possible below “Application Term”
Multi-domain Thesaurus
(shared, stable, multi-lingual)
Semantic links

System Specific
Taxonomy
(dynamic, selected language)
Multi-domain Thesaurus

Application Terms can be related to themes, domains, information, and measures.

Themes:
- health
- environment
- climate

Domains:
- deterioration of environment
- natural environment

Information:
- marine deterioration
- terrestrial environment
- marine environment

Measures:
- water quality
- oil pollution
- alga bloom
- water turbidity
- water transparency
- Secchi depth
- oil spill drift forecast
- oil spill surveillance
- oil spill monitoring
- alga bloom location / extent
- wave height
- wave period
- alga bloom map
- alga bloom monitoring
- waves

Multi-domain Thesaurus
<table>
<thead>
<tr>
<th>Term Type</th>
<th>Term</th>
<th>Application Domain</th>
<th>Description</th>
<th>Synonyms</th>
<th>Related terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>alga bloom</td>
<td>marine deterioration</td>
<td>An alga bloom is a rapid increase in the population …</td>
<td>harmful alga bloom, marine bloom, water bloom, …</td>
<td>eutrophication, …</td>
</tr>
<tr>
<td>Application</td>
<td>alga bloom</td>
<td>marine environment</td>
<td>“</td>
<td>“</td>
<td>“</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>
Proposed approach

1. Collect outline information on existing ontologies and taxonomies associated with Earth observation, and provide an overview.

2. Compare and analyze the ontologies and taxonomies.
   - Scope, Objectives, Language?
   - Overlaps? Relationships? Maintenance?
   - Models? (format, logical structure etc.)

3. Populate Best Practice Wiki on Ontologies/Taxonomies as referenceable basis for GEOSS.

4. Assess the possibility of a “Standard” Ontology/Taxonomy for GEOSS.

5. Assist in the deployment of the referenceable ontologies and taxonomies in Best Practice Wiki for Earth observation in populating the Registries.

6. Assess how to use the ontologies and taxonomies for discovery composition and access in the frame of the GEOSS architecture.
Developing the support tool for generating metadata

Data provider

Manual input

Metadata registration interface

Storing to database

Publishing

Document

User

Automatic extraction

Datasets

Reference

Thesauri Ontologies Taxonomies

Best Practice Wiki
Proposed approach

1. Collect outline information on existing ontologies and taxonomies associated with Earth observation, and provide an overview.

2. Compare and analyze the ontologies and taxonomies.
   + Scope, Objectives, Language?
   + Overlaps? Relationships? Maintenance?
   + Models? (format, logical structure etc.)

3. Populate Best Practice Wiki on Ontologies/Taxonomies as referenceable basis for GEOSS.

4. Assess the possibility of a “Standard” Ontology/Taxonomy for GEOSS.

5. Assist in the deployment of the referenceable ontologies and taxonomies in Best Practice Wiki for Earth observation in populating the Registries.

6. Assess how to use the ontologies and taxonomies for discovery composition and access in the frame of the GEOSS architecture.
Multi-domain Search Tool

(http://dimeola.esrin.esa.int/OTE/navigateInfoDomain)

Multi-domain Thesaurus and Vocabulary

Requires Java 1.6

Link to 2D Navigator
Reverse Dictionary
to help find exact entry words for search
Possible use case of ontology/taxonomy registry for GCI
Draft Roadmap

1. Collect outline information (inventory development) on existing ontologies and taxonomies associated with Earth observation, and provide an overview. ➔ Dec. 2009

2. Compare and analyze the ontologies and taxonomies. ➔ Mar. 2010

3. Populate Best Practice Wiki on Ontologies/Taxonomies as referenceable basis for GEOSS. ➔ Mar. 2010

3. Assess the possibility of “Standard” Ontology/Taxonomies for GEOSS. ➔ After the end of AIP-3?

4. Assist in the deployment of the referenceable ontologies and taxonomies in populating the registries. ➔ AIP-3?

5. Assess how to use the ontologies and taxonomies for discovery composition and access in the frame of the GEOSS architecture. ➔ AIP-3?
Immediate actions

• Monthly telecon
• Start the development of Wiki-based inventory
  – Invite volunteers by circulating CFP among GEO members. (CEOS, WMO etc.)
  – Collect information on existing ontology/taxonomy.
  – Determine which to be invented.
  – Develop the inventory.

Earth Observation ontology/taxonomy should be prioritized.
Geospatial domain related concepts

- **Gazetteer**
  - Digital gazetteer is the dictionary of types, names and locations.
  - Focuses on database equipment.

- **Geospatial ontology**
  - Geospatial ontology is the knowledge resources about geospatial features, place names or spatial relations.
  - Focuses conceptual modeling

- **Geocoding**
  - Geocoding is the act of transforming descriptive text into a valid spatial representation.
  - Focuses on translation process.

<table>
<thead>
<tr>
<th>Feature Types</th>
<th>Name</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td>Melbourne</td>
<td>37° 49’ 14” S, 144° 57’ 41” E</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query</th>
<th>Gazetteer</th>
<th>Geocoding</th>
</tr>
</thead>
</table>

- **Resources**
  - Features
  - Place names
  - Spatial relations
Quality Criteria of Digital Gazetteer

- The quality element in evaluating each digital gazetteers.
  - 1) Availability
    - Use constraint or restriction for digital gazetteer (ex. right, security, policy)
  - 2) Scope
    - Target scope on earth surface. Regional/ National/ Worldwide level?
  - 3) Completeness
    - Degrees on covering various kinds of data within target scope
  - 4) Currency
    - Handling famous, familiar name? Managing update and change?
  - 5) Accuracy
    - Degrees of correctness for name or accuracy for location.
  - 6) Granularity
    - Degrees of resolution or scale
  - 7) Richness of annotation/description/information
    - Association with database.