Harmonising Geospatial Thesauri: Groundwork for semantics-aware interoperability

Cristiano Fugazza - cristiano.fugazza@jrc.ec.europa.eu
Presenting: Lorenzino Vaccari - lorenzino.vaccari@jrc.ec.europa.eu

Full presentation: http://prezi.com/om0n9zi2zs0h/harmonising-geospatial-thesauri/
The problem:

How to retrieve a resource annotated by the keyword "elevation" when the search string entered by the user is "海拔" ("altitude").
Two main issues:
Multilingualism: The search string cannot be matched against metadata because it is expressed in a language which is different from that used for annotation.

海拔 ≠ altitude
Semantic heterogeneity: The search term may be different (e.g., more general, more specific, synonym of) from that used for annotation

\[ \text{altitude} \neq \text{elevation} \]
The solution:

In either the annotation and/or the discovery phase, pick terms from **controlled vocabularies**.
Widely acknowledged examples of controlled vocabularies in the Simple Knowledge Organisation System (SKOS) format, a.k.a. thesauri, already exist.
We currently host:

Reference thesauri:

- GEneral Multilingual Environmental Thesaurus (GEMET)
- INSPIRE Feature Concept Dictionary (IFCD)
- INSPIRE Glossary
- INSPIRE Themes
- GEOSS Societal Benefit Areas (SBA)
- ISO 19119 Geographic Services Taxonomy
- GEOSS Earth Observation Vocabulary
- GCMD Science Keywords

We currently host:

Thematic thesauri:

- GEOSS AIP-3 Water Ontology
- EuroGEOSS Drought Vocabulary
And, most importantly:

**Mappings between thesauri:**

- GEOSS Societal Benefit Areas ➔ GEneral Multilingual Environmental Thesaurus (GEMET)
- GEOSS Societal Benefit Areas ➔ INSPIRE Themes
- GEOSS Societal Benefit Areas ➔ EuroGEOSS Drought Vocabulary
- GEOSS Societal Benefit Areas ➔ GEOSS Earth Observation Vocabulary
- GEOSS Earth Observation Vocabulary ➔ GEneral Multilingual Environmental Thesaurus (GEMET)
- GEOSS Earth Observation Vocabulary ➔ INSPIRE Themes
- GEOSS Earth Observation Vocabulary ➔ GCMD Science Keywords
- INSPIRE Themes ➔ GEneral Multilingual Environmental Thesaurus (GEMET)
Step #1: Match the user's search string against terms' text representation

```sql
01 SELECT DISTINCT ?term ?lab ?thesaurus
02 WHERE
03 {
05 }
06 UNION
07 {
08 ?term skos:exactMatch ?term OPTIONAL (TRANSITIVE, T_MAX 1).
10 }
14 }
15 FILTER (?label = "altitude")
16 FILTER (?thesaurus = "GEMET, version 2.4")
17 FILTER (?label = "en")
18 FILTER (?thesaurus = "en")
19 }
20 )
21 \n22 }
23 }
24 }
```

The search string is matched against equivalent terms which may not have translations in the user's language.

The regular expression matches the user's search string in a case-insensitive way and considering leading and trailing punctuation.

Result: The search string is translated into language-neutral identifiers

<table>
<thead>
<tr>
<th>term</th>
<th>lab</th>
<th>thesaurus</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.eionet.europa.eu/gemet/concept/10140">http://www.eionet.europa.eu/gemet/concept/10140</a></td>
<td>&quot;altitude&quot;@en</td>
<td>&quot;GEMET, version 2.4&quot;@en</td>
</tr>
</tbody>
</table>
Resource discovery:

Step #2: Expand the results of Step #1 according to semantic links

01 SELECT DISTINCT ?term ?lab ?thesaurus ?prop
02 WHERE
03 {
04   {
06   }
07   UNION
08   {
10   }
11   }
12  OPTIONAL
13  {
15  }
16  FILTER ( LANG(?lab) = "en"
17  )
18 }
19
Result: The set of terms considered by the query is increased

<table>
<thead>
<tr>
<th>term</th>
<th>lab</th>
<th>thesaurus</th>
<th>property</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.eionet.europa.eu/gemet/concept/6033">http://www.eionet.europa.eu/gemet/concept/6033</a></td>
<td>&quot;parameter&quot;@en</td>
<td>&quot;GEMET, version 2.4&quot;@en</td>
<td>http://.../core#broader</td>
</tr>
<tr>
<td><a href="http://www.earthobservations.org/.../elevation@en">http://www.earthobservations.org/.../elevation@en</a></td>
<td>&quot;Elevation&quot;@en</td>
<td>&quot;GEOSS - Earth Observation Vocabulary, version 1.0&quot;</td>
<td>http://.../core#closeMatch</td>
</tr>
</tbody>
</table>
Step #3: Translate results back to multiple languages

```sql
01 SELECT DISTINCT ?lab_en ?lab_nl ...
02 WHERE
03 {
04   OPTIONAL
05   {
07   }
08   FILTER {
09     LANG(?lab_en) = "en"
10   }
11   OPTIONAL
12   {
14   }
15   FILTER {
16     LANG(?lab_nl) = "nl"
17   }
18 }
```

Result: Multiple, language-dependent queries can be issued to catalogue services

<table>
<thead>
<tr>
<th>lab_en</th>
<th>lab_nl</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;altitude&quot;@en</td>
<td>&quot;hoopte&quot;@nl</td>
<td>...</td>
</tr>
<tr>
<td>&quot;Elevation&quot;@en</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Resource discovery:

Selective browsing of thesauri allows for the selection of individual terms to be included in the query.
Resource annotation:

Step #1: Select a thesaurus where to choose from

```
01 SELECT DISTINCT ?voc ?lab ?date
02 WHERE {
03   {
04     ?voc skos:prefLabel ?lab .
05     ?voc skos:prefix skos:ConceptScheme.
06     OPTIONAL { ?voc skos:notation ?date .}
07     } FILTER ( LANG(?lab) = "en")
08   }
09 }
10 ORDER BY ASC(?lab)
```
Resource annotation:

Step #2: List topmost terms in the thesaurus

```sql
SELECT DISTINCT ?con ?lab ?alt ?td
WHERE {
  ?con rdfs:label ?lab.
  Optional { ?con skos:notation ?an .}
  Optional { ?con skos:broader  ?b .}
  OPTIONAL { ?con skos:notation ?an .}
  OPTIONAL { ?con skos:broader  ?b .}
  OPTIONAL { ?con skos:preferredLabel ?pl .}
  FILTER ( ?lab = "en" )
  ORDER BY ASC( ?lab )
}
```

A non-empty value in this column means that the term has more specific ones.

<table>
<thead>
<tr>
<th>?con</th>
<th>?lab</th>
<th>?alt</th>
<th>?td</th>
</tr>
</thead>
</table>
Resource annotation:

Step #3: List more specific terms w.r.t. a give one

```
SELECT DISTINCT ?con ?lab ?d
WHERE {
  { 
    ?con rdf:type ?lab.
    ?con skos:broader <http://grand.gis.nas.edu/rdfs/agriculture/>,
    OPTIONAL { 
      ?con skos:notation ?n.
    }
  }
  OPTIONAL { 
    ?con skos:prefLabel ?n.
    FILTER ( LANG(?n) = "en")
    FILTER ( LANG(?lab) = "en")
  }
  UNION
  { 
    # DO THE SAME FOR COLLECTIONS
  }
} ORDER BY ASC(?d) ASC(?lab)
```
Resource annotation:
Step #4: Select terms for inclusion in metadata

Either the text representation of terms and the unique identifier (URI) can be added to metadata descriptions of resources.
Conclusions:

- The SKOS data model can be used for representing and harmonising thesauri for the geospatial domain
- RDF triple stores allow for deployment of thesauri and applications can access these through HTTP
- SPARQL queries can be tailored so as to suit a variety of diverse use cases

Currently, the vocabulary service outlined in this presentation is used for:
- Query expansion in the EuroGEOSS discovery broker
- Annotation of the resources in the EuroGEOSS INSPIRE metadata editor
- Semantic enablement of the GEOSS Common Infrastructure

The next step is:
- Indexing of resources in the INSPIRE Geoportal
Harmonising Geospatial Thesauri:
Groundwork for semantics-aware interoperability


Cristiano Fugazza - cristiano.fugazza@jrc.ec.europa.eu
Presenting: Lorenzino Vaccari - lorenzino.vaccari@jrc.ec.europa.eu