Comparing Vocabularies for Representing Geographical Features and Their Geometry

Ghislain Atemezing / @gatemezing
Raphaël Troncy / @rtroncy
Goal and Research Questions

- **Goal:** Provide recommendations for the French IGN in exposing their GIS database in the Linked Data world
- **Study of geo vocabularies in the Web of Data**
  - LOD Cloud review
  - Who are the GeoData providers?
  - How features are generally modeled?
  - How geometry is generally modeled?
  - Illustrative scenario
- **Align vocabularies when necessary**
- **Compare Triple Stores with geospatial indexing**
GeoData: why it matters?

- “80% of needs for decisions from public authorities have a geospatial component”. (Philippe Grelot, IGN-France)

Photo by Vilavelosa on flickr.com
## Feature and Geometry

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Object = (Geographical) Feature</td>
<td>An abstract representation of a real-world phenomenon related to a specific location or geographical area. It should be noted that the term has a different meaning in the ISO 19100 series. It is also synonymous with &quot;(geographic) feature&quot; as used in the ISO 19100 series.</td>
<td>[INSPIRE Directive] Item 67</td>
</tr>
<tr>
<td>Feature</td>
<td>A geographical feature, capable of holding spatial relations.</td>
<td>NeoGeo Vocab</td>
</tr>
<tr>
<td>Geometry</td>
<td>A top-level geometry type. This class is equivalent to the UML class GM_Object defined in ISO 19107, and it is superclass of all geometry types.</td>
<td>GeoSPARQL [OGC]</td>
</tr>
</tbody>
</table>
GeoData on the LOD Cloud

http://lod-cloud.net/state


31 datasets
19.43% triples
Where are Geo-Linked Data?

- DBpedia
- GeoNames
- LinkedGeodata (OSM)
- Freebase (Google)
- Ordnance Survey (UK)
- GeoLinkedData (ES)
- GADM-RDF
- NUTS-RDF
- data.ign.fr (FR)

<table>
<thead>
<tr>
<th>Provider</th>
<th>#Triples</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBpedia</td>
<td>727,232 triples</td>
</tr>
<tr>
<td>GeoNames</td>
<td>5,240,032 « features »</td>
</tr>
<tr>
<td>LinkedGeoData</td>
<td>60,356,364 triples</td>
</tr>
<tr>
<td>Ordnance Survey</td>
<td>6,295 triples</td>
</tr>
<tr>
<td>Freebase</td>
<td>8,5 MB (tsv fichiers)</td>
</tr>
<tr>
<td>GeoLinkedData.es</td>
<td>101,018 triples</td>
</tr>
<tr>
<td>Projet GADM</td>
<td>682,605 triples</td>
</tr>
<tr>
<td>Projet NUTS</td>
<td>316,238 triples</td>
</tr>
</tbody>
</table>
## Vocabularies for Space and Geography

### Vocabularies member list:

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Namespace</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>ad</td>
<td><a href="http://schemas.talis.com/2005/address/schema#">http://schemas.talis.com/2005/address/schema#</a></td>
<td>Address Schema</td>
</tr>
<tr>
<td>geo</td>
<td><a href="http://www.w3.org/2003/01/geo/wgs84_pos#">http://www.w3.org/2003/01/geo/wgs84_pos#</a></td>
<td>WGS84 Geo Positioning</td>
</tr>
<tr>
<td>geod</td>
<td><a href="http://vocab.lenka.no/geo-deling#">http://vocab.lenka.no/geo-deling#</a></td>
<td>Land classification in Norway</td>
</tr>
<tr>
<td>geof</td>
<td><a href="http://www.mindswap.org/2003/owl/geo/geoFeatures20040307.owl#">http://www.mindswap.org/2003/owl/geo/geoFeatures20040307.owl#</a></td>
<td>Geo Features</td>
</tr>
<tr>
<td>geop</td>
<td><a href="http://aims.fao.org/aos/geopolitical.owl#">http://aims.fao.org/aos/geopolitical.owl#</a></td>
<td>FAO Geopolitical Ontology</td>
</tr>
<tr>
<td>geos</td>
<td><a href="http://www.telegraphis.net/ontolog/geography/geography#">http://www.telegraphis.net/ontolog/geography/geography#</a></td>
<td>Geographies Ontology</td>
</tr>
<tr>
<td>gn</td>
<td><a href="http://www.geonames.org/ontology#">http://www.geonames.org/ontology#</a></td>
<td>The Geonames Ontology</td>
</tr>
<tr>
<td>gnm</td>
<td><a href="http://www.geonames.org/ontology/mappings/">http://www.geonames.org/ontology/mappings/</a></td>
<td>Geonames mappings</td>
</tr>
<tr>
<td>lgdo</td>
<td><a href="http://linkedgeodata.org/ontology/">http://linkedgeodata.org/ontology/</a></td>
<td>LinkedGeoData ontology</td>
</tr>
<tr>
<td>loc</td>
<td><a href="http://purl.org/ci/ciinfraestructuras/localizacion#">http://purl.org/ci/ciinfraestructuras/localizacion#</a></td>
<td>Vocabulario de Localizaciones</td>
</tr>
<tr>
<td>ngeo</td>
<td><a href="http://geovocab.org/geometry#">http://geovocab.org/geometry#</a></td>
<td>NeoGeo Geometry Ontology</td>
</tr>
<tr>
<td>osadm</td>
<td><a href="http://data.ordnancesurvey.co.uk/ontology/osadmingeo/">http://data.ordnancesurvey.co.uk/ontology/osadmingeo/</a></td>
<td>The administrative geography and civil voting area ontology</td>
</tr>
<tr>
<td>osr</td>
<td><a href="http://purl.org/ontomedia/core/space#">http://purl.org/ontomedia/core/space#</a></td>
<td>OntoMedia Space Representation</td>
</tr>
<tr>
<td>osspr</td>
<td><a href="http://data.ordnancesurvey.co.uk/ontology/spatialrelations/">http://data.ordnancesurvey.co.uk/ontology/spatialrelations/</a></td>
<td>Spatial Relations Ontology</td>
</tr>
<tr>
<td>osstop</td>
<td><a href="http://www.ordnancesurvey.co.uk/ontology/Topography/v0.1/Topography.owl#">http://www.ordnancesurvey.co.uk/ontology/Topography/v0.1/Topography.owl#</a></td>
<td>Ordinance Survey Topography Ontology</td>
</tr>
<tr>
<td>place</td>
<td><a href="http://purl.org/ontology/places#">http://purl.org/ontology/places#</a></td>
<td>The Places Ontology</td>
</tr>
<tr>
<td>spatial</td>
<td><a href="http://geovocab.org/spatial#">http://geovocab.org/spatial#</a></td>
<td>NeoGeo Spatial Ontology</td>
</tr>
</tbody>
</table>

Only 5 vocabs are reused: **W3C Geo (21 datasets), OS spatialrelations (10 datasets), Geonames (5 datasets), UK administrative (3 datasets) and NeoGeo (3 datasets)**
Vocabularies for Modeling Features (1/2)

- Authority list of terms (e.g. Foursquare)
  - Less structured
  - Represent categories of Points of Interest (POIs)
  - Typically, one type as an API answer
  - Need to express the semantics of the terms

- SKOS Categories (e.g. GeoNames)
  - Classes are \texttt{skos:conceptScheme}
  - Codes are \texttt{skos:Concept}
  - Few classes … BUT many codes
Vocabularies for Modeling Features (2/2)

- **Domain specific ontologies**
  - One ontology per subdomain
    - (transport, administrative unit, hydrography, etc.)
  - Interconnected ontologies
    - (by explicit semantic e.g. `owl:imports`)
  - UK (OS) – ES (GeoLinkedData)

- **Some richer ontologies created by (semi-)automatic tools / NLP**
  - Deeper taxonomy to structure the ontology
  - **LinkedGeoData**: 16 high-level classes, 1294 classes
  - **GeOnto**: 2 high-level classes, 783 classes in total
Modeling Geometry

- **Point (lat/long)**
  - WGS 84 vocabulary described by W3C

- **Rectangle (“bounding box”)**
  - Geopolitical Vocabulary (FAO)

- **Points in a List**
  - Sequence of points (LinkedGeoData)
  - An object is “formedBy” a ListOfPoints (GeoLinkedData.es)

- **Literals (GML datatype in RDF)**
  - Ordnance Survey (UK)

- **More structured representation of complex geometry**
  - NeoGeo Vocabulary (GeoVocamp), [http://geovocab.org/](http://geovocab.org/)
Scenario: 7th Arrondissement of Paris

The 7th arrondissement of Paris is one of the 20 arrondissements (administrative districts) of the capital city of France. It includes some of Paris's major historical sites, such as the Luxembourg Gardens, the Montparnasse Tower, and the Place de la Bastille. The arrondissement is also home to many foreign diplomatic missions, including the United States Embassy.

Situated in the Left Bank of the Seine River, the 7th arrondissement is known for its rich history and cultural significance. It was the center of Jewish culture in Paris until the 20th century, and today it is home to many cultural institutions, including the French National Assembly and numerous government ministries.

The arrondissement has been home to French upper class since the 17th century, when it became the new residence of French highest nobility. The district has been referred to as the "golden age of the 7th arrondissement". It is considered one of the most picturesque districts in Paris.
7th Arrondissement in DBpedia (a gml_Feature)

dbpedia:7th_arrondissement_of_Paris a gml:_Feature ;
   (gml IS NOT an ontology with OWL-flavour )
   a <http://dbpedia.org/class/yago/1900SummerOlympicVenuEs>
(Yago Class)

rdfs:label "巴黎第七區"@zh; (14 different languages)
dbpprop:commune "Paris" ;
dbpprop:département dbpedia:Paris ;
dbpprop:région dbpedia:Île-de-France_(region) ;
grs:point "48.85916666666667 2.312777777777778" ;
geo:geometry "POINT(2.31278 48.8592)" ; (fake property!)
geo:lat "48.859165"^^xsd:float;
geo:long "2.312778"^^xsd:float.
7th Arrondissement in GeoNames (a A.ADM4)

gnr:6618613 a gn:Feature ; gn:name "Paris 07";
  gn:alternateName "7ème arrondissement";
  gn:featureClass gn:A [
    a skos:ConceptScheme ;
    rdfs:comment "country, state, region ..."@en .
  ] ;

  gn:featureColde gn:A.ADM4 [
    a skos:Concept ;
    rdfs:comment
    "a subdivision of a third-order administrative division"@en .
  ];

  gn:countryCode "FR";
  gn:population "57410";
  geo:lat "48.8565";
  geo:long "2.321".

2012/11/12
7th Arrondissement in LGD (a “Suburb”)

```
lgd:node248177663 a lgdo:Suburb ;
  rdfs:label "7th Arrondissement"@en , "7e Arrondissement" ;
  lgdo:contributor lgd:user13442 ;
  <http://linkedgeodata.org/ontology/ref%3AINSEE> 75107 ;
  lgdp:alt_name "VIIe Arrondissement" ;
  georss:point "48.8570281 2.3201953" ;
  geo:lat 48.8570281 ;
  geo:long 2.3201953 .
```
Different databases:
- BD ORTHO
- BD PARCELLEindre
- POINT ADRESSE
- BD ALTI 25m
- BD TOPO

Data in LAMBERT93 or RGF93

Q: "Give me all the bridges in a radius of 2km from the "Eiffel Tower"?
A: Not straightforward
Modeling Features in France (GeOnto)

- Ontology for geographic objects (POI)
  - Output of a French (ANR) research project
  - Obtained from NLP tools

- Classes in French
  - `rdfs:labels` in FR & EN
  - No `rdfs:comments`
  - Few `owl:ObjectProperty`
  - 783 classes

- Overlap with other vocabbs
  - Need for alignment
Alignment Methodology

- Alignment of GeOnto with 4 ontologies and 2 more simple taxonomies
  - LGD, DBpedia, Schema.org, GeoNames
  - Foursquare, Google Places

- Goal: finding `owl:equivalentClass`
  - Tool: Silk
  - Metrics: LevenshteinDistance, Jaro
  - Labels: @en des classes
  - Aggregation Function: Mean

- Manual validation
  - For `rdfs:subClassOf`
  - Specific alignments with GeoNames codes
Alignment Process with GeoNames

\[ \text{geOnto:AGeoConcept} \text{ a}\]
\[ \text{owl:Class;} \]
\[ \text{rdfs:label } \text{“a label”} @\text{en;} \]
\[ \text{rdfs:subClassOf gn:Feature;} \]
\[ \text{owl:equivalentClass} \]
\[ \text{[a owl:Restriction;} \]
\[ \text{owl:onProperty} \]
\[ \text{gn:featureCode;} \]
\[ \text{owl:hasValue gn:CODE.} \]

- Look for \textit{skos} codes that matches GeOnto classes
- Verify the links <70%
- Generate « sameAs » links

- Use SPARQL
« \textit{Construct} » to generate a new graph.

- Export the rdf file
Results/Evaluation

<table>
<thead>
<tr>
<th>Vocab/taxonomies</th>
<th>#Classes</th>
<th>#Classes aligned</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGD</td>
<td>owl:Class: 1294</td>
<td>178</td>
</tr>
<tr>
<td>DBpedia</td>
<td>owl:Class: 366</td>
<td>42</td>
</tr>
<tr>
<td>Schema.org</td>
<td>owl:Class: 296</td>
<td>52</td>
</tr>
<tr>
<td>GeoNames</td>
<td>skos:Concept: 699</td>
<td>287</td>
</tr>
<tr>
<td>Foursquare</td>
<td>359</td>
<td>46</td>
</tr>
<tr>
<td>Google Place</td>
<td>126</td>
<td>41</td>
</tr>
<tr>
<td>bdtopo</td>
<td>owl:Class: 237</td>
<td>153</td>
</tr>
</tbody>
</table>

- High precisions > 80%
- BUT P(Schema.org) = 50%.
  - Possible reasons: GeOnto entities are more specific to France
  - Fine grain details for entities in Schema.org
<table>
<thead>
<tr>
<th>Geo-vocabulary</th>
<th>Topological Functions</th>
<th>GeoSPARQL Requirements</th>
<th>Standard Followed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordnance Survey Spatial</td>
<td>equal, greater, less, touches, within, contains</td>
<td>Part of Req 4</td>
<td>OpenGIS Simple Feature</td>
</tr>
<tr>
<td>Ordnance Survey Topography</td>
<td>contains, isContainedIn</td>
<td>Very small part of Req 4</td>
<td>OpenGIS Simple Feature</td>
</tr>
<tr>
<td>Place Ontology</td>
<td>in, overlaps, bounded_by</td>
<td>Small part of Req 4</td>
<td>N/A</td>
</tr>
<tr>
<td>NeoGeo Spatial</td>
<td>All RCC8 relations</td>
<td>Part of Req 3; Req 6</td>
<td>Region Connection Calculus (RCC)</td>
</tr>
<tr>
<td>NeoGeo Geometry</td>
<td>—</td>
<td>Req 10 - 14</td>
<td>N/A</td>
</tr>
<tr>
<td>FAO Geopolitical</td>
<td>isInGroup, hasBorderWith</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>OntoMedia Space</td>
<td>adjacent-below, adjacent-above, orbit-around, is_boundary-of, has-boundary</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

NeoGeo (Spatial) and OS Spatial have integrated in their modeling partial or full aspect of topological functions of GeoSPARQL.
### Triple Stores and geospatial indexing

<table>
<thead>
<tr>
<th>Triple store</th>
<th>WKT-compliance</th>
<th>GML-compliance</th>
<th>Geometry supported</th>
<th>Geospatial Functions</th>
<th>GeoVocab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtuoso</td>
<td>Yes</td>
<td>Yes</td>
<td>Point</td>
<td>13 functions</td>
<td>W3C Geo + Typed Literal</td>
</tr>
<tr>
<td>Allegro-Graph</td>
<td>-</td>
<td>-</td>
<td>Point</td>
<td>3 functions</td>
<td>“strip” mapping data</td>
</tr>
<tr>
<td>OWLIM-SE</td>
<td>-</td>
<td>-</td>
<td>Point</td>
<td>4 functions</td>
<td>W3C Geo</td>
</tr>
<tr>
<td>Open Sahara</td>
<td>Yes</td>
<td>Yes</td>
<td>Point, Line, Polygons</td>
<td>23 functions</td>
<td>Typed Literal</td>
</tr>
<tr>
<td>Parliament</td>
<td>Yes</td>
<td>Yes</td>
<td>Point, Line, Polygons</td>
<td>23 functions</td>
<td>GeoSPARQL vocabulary</td>
</tr>
</tbody>
</table>

Open Sahara, Parliament and Virtuoso are good choices because they integrate many Geospatial Functions.
Some Recommendations

- **Complex Geometry Coverage**
  - Need to publish more data with complex geometries
  - Select suitable ontologies (e.g: NeoGeo) or GeoSPARQL

- **Features MUST be connected to Geometry**
  - Sometimes it may requires two namespaces

- **Serialization and Triple Stores**
  - Provide serialization in other GIS formats (GML, WKT, KML, etc.)
  - Store geodata in a triple store with many topological functions implemented (e.g: Open Sahara, Parliament, Virtuoso)

- **Literal vs Structured Representation**
  - Use of structured representation for complex geometry
  - This covers some of the Use Cases at IGN
**Conclusion**

- **Studied geo vocabularies in the Web of Data**
  - Multiplicity, attempt of comparison
  - Alignment needed, starting from a new ontology

- **Presented steps tailored for the French IGN**
  - Publishing following linked data principles …
  - … including complex geometry

- **Outline some recommendations**
  - When publishing data with complex geometry on the web
  - Useful for any Geodata provider having similar requirements than IGN
Future Work

- Publish a new version of GeOnto ontology
  - Following the Best Practices on the LOD
  - Reusing NeoGeo Vocabulary
  - Use of W3C Geo for representing points

- « Lift » raw data in RDF
  - Using GeOnto and external vocabularies
  - Store graph in Virtuoso + IndexingSail service

- Continue mappings and alignments
  - Schema.org, Foursquare, Google Place
  - GeoSPARQL vocabulary
  - Mappings at data level
Thanks for your attention!

Questions?
French IGN & Open Data Initiative

- Provider of the data.gouv.fr portal
- 21 datasets in SHAPE files

- Want to publish their data in 5 stars
- Data.ign.fr (experimental version)
- Towards IGN LD with complex geometries