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)**

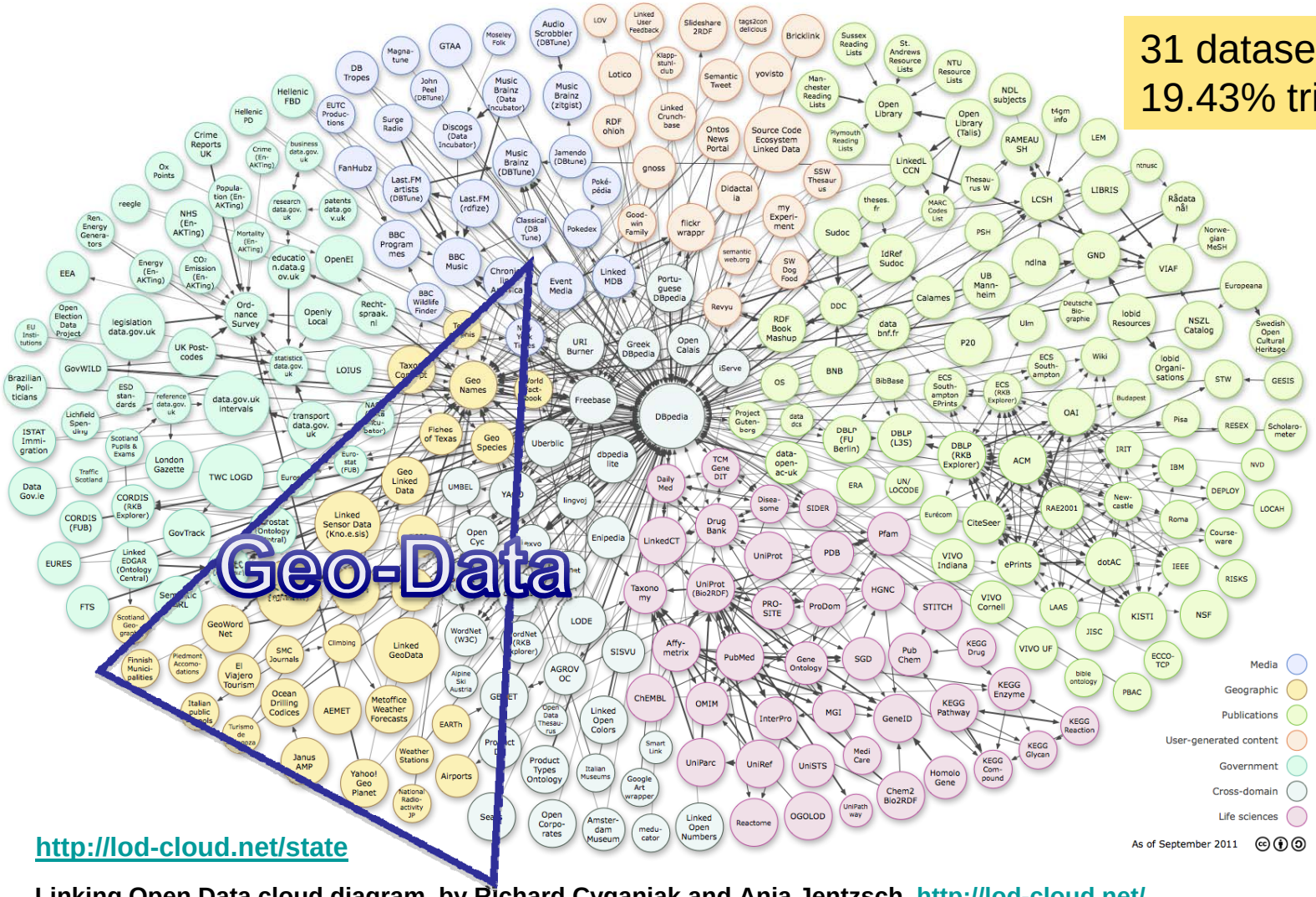
Photo by Vilavelosa on flickr.com

Feature and Geometry

Term	Definition	Source
Spatial Object = (Geographical) Feature	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 "(geographic) feature" as used in the ISO 19100 series.	[INSPIRE Directive] Item 67
Feature	A geographical feature, capable of holding spatial relations.	NeoGeo Vocab
Geometry	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.	GeoSPARQL [OGC]

GeoData on the LOD Cloud

31 datasets
19.43% triples



http://lod-cloud.net/state

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

EURECOM
Sophia Antipolis

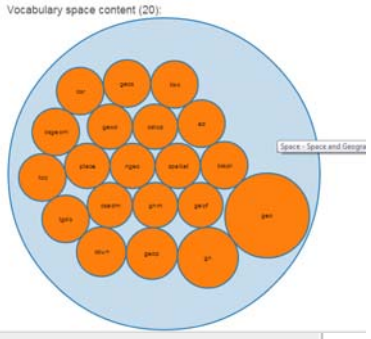
Where are Geo-Linked Data?

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

Provider	#Triples
DBpedia	727 232 triples
GeoNames	5 240 032 (« features »)
LinkedGeoData	60 356 364 triples
Ordnance Survey	6 295 triples
Freebase	8,5 MB (tsv fichiers)
GeoLinkedData.es	101 018 triples
Projet GADM	682 605 triples
Projet NUTS	316 238 triples

Vocabularies for Space and Geography

Vocabularies member list: http://lov.okfn.org/dataset/lov/details/vocabularySpace_Space.html



Prefix	Namespace	Title
ad	http://schemas.talis.com/2005/address/schema#	Address Schema
coun	http://www.daml.org/2001/09/countries/iso-3166-ont#	ISO 3166 Country Codes
geo	http://www.w3.org/2003/01/geo/wgs84_pos#	WGS84 Geo Positioning
geod	http://vocab.lenka.no/geo-deling#	Land classification in Norway
geof	http://www.mindswap.org/2003/owl/geo/geoFeatures20040307.owl#	Geo Features
geop	http://aims.fao.org/aos/geopolitical.owl#	FAO Geopolitical Ontology
geos	http://www.telegraphis.net/ontology/geography/geography#	Geographis Ontology
gn	http://www.geonames.org/ontology#	The Geonames Ontology
gnm	http://www.geonames.org/ontology/mappings/	Geonames mappings
lgdo	http://linkedgeodata.org/ontology/	LinkedGeoData ontology
loc	http://purl.org/ctic/infraestructuras/localizacion#	Vocabulario de Localizaciones
ngeo	http://geovocab.org/geometry#	NeoGeo Geometry Ontology
osadm	http://data.ordnancesurvey.co.uk/ontology/admingeo/	The administrative geography and civil voting area ontology
osr	http://purl.org/ontomedia/core/space#	OntoMedia Space Representation
osspr	http://data.ordnancesurvey.co.uk/ontology/spatialrelations/	Spatial Relations Ontology
ostop	http://www.ordnancesurvey.co.uk/ontology/Topography/v0.1/Topography.owl#	Ordnance Survey Topography Ontology
place	http://purl.org/ontology/places#	The Places Ontology
spatial	http://geovocab.org/spatial#	NeoGeo Spatial Ontology

Only 5 vocabs are reused: **W3C Geo (21 datasets), OS spatialrelations (10 datasets), Geonames (5 datasets), UK administrative (3 datasets) and NeoGeo (3 datasets)**

EURECOM Sophia Antipolis

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 `skos:conceptScheme`
  - Codes are `skos:Concept`
  - Few classes … BUT many codes

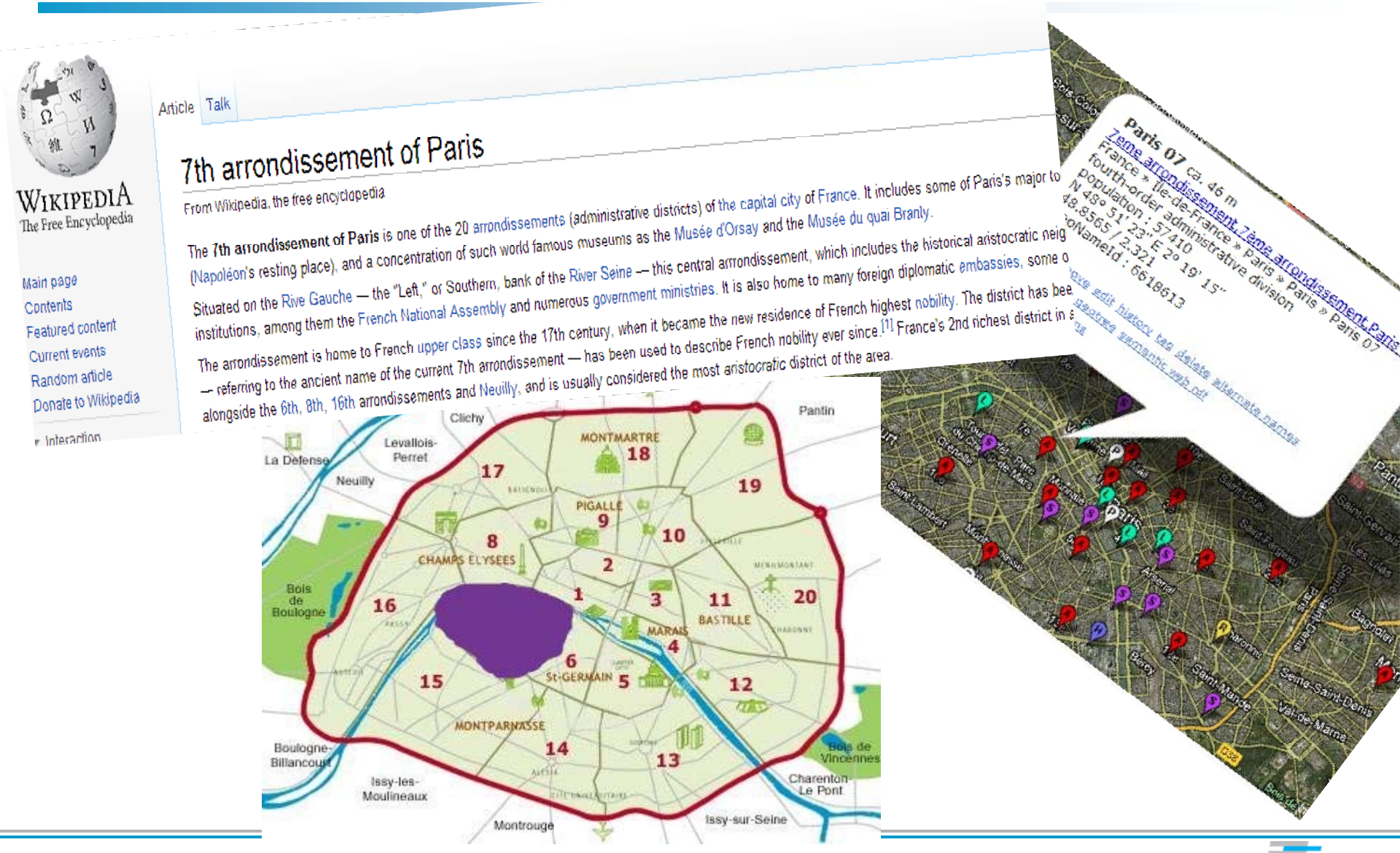
Vocabularies for Modeling Features (2/2)

- **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/**

Scenario: 7th Arrondissement of Paris



EURECOM
Sophia Antipolis

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 "48.8591666666667 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".
```

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 .
```

French IGN

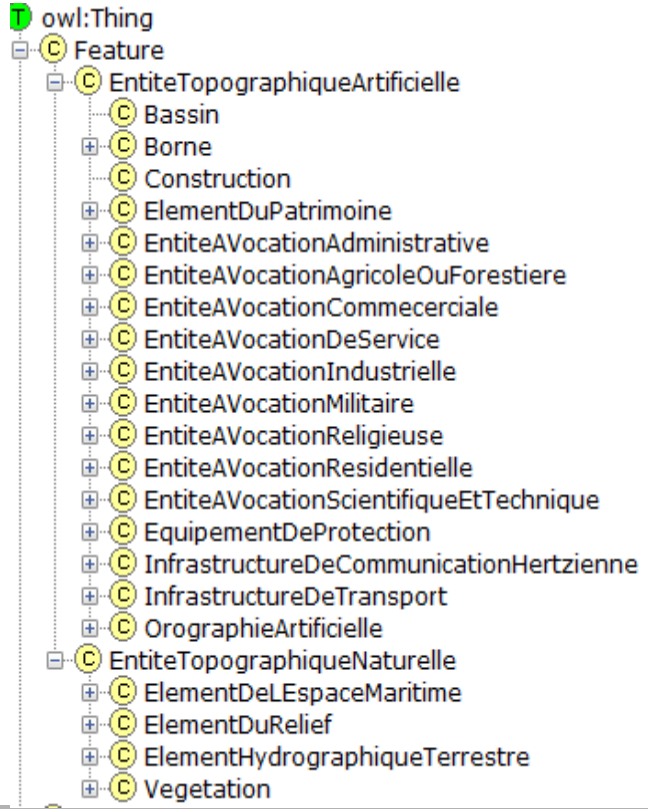
« ..describes the French national territory and the occupation of its land, elaborates and updates perpetual inventory of the forest resources »

- ✓ Different databases:
  BD ORTHO, BD PARCELLAIRE, POINT ADRESSE, BD ALTI 25m, BD TOPO; etc.
- ✓ 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 vocabs**
  - Need for alignment

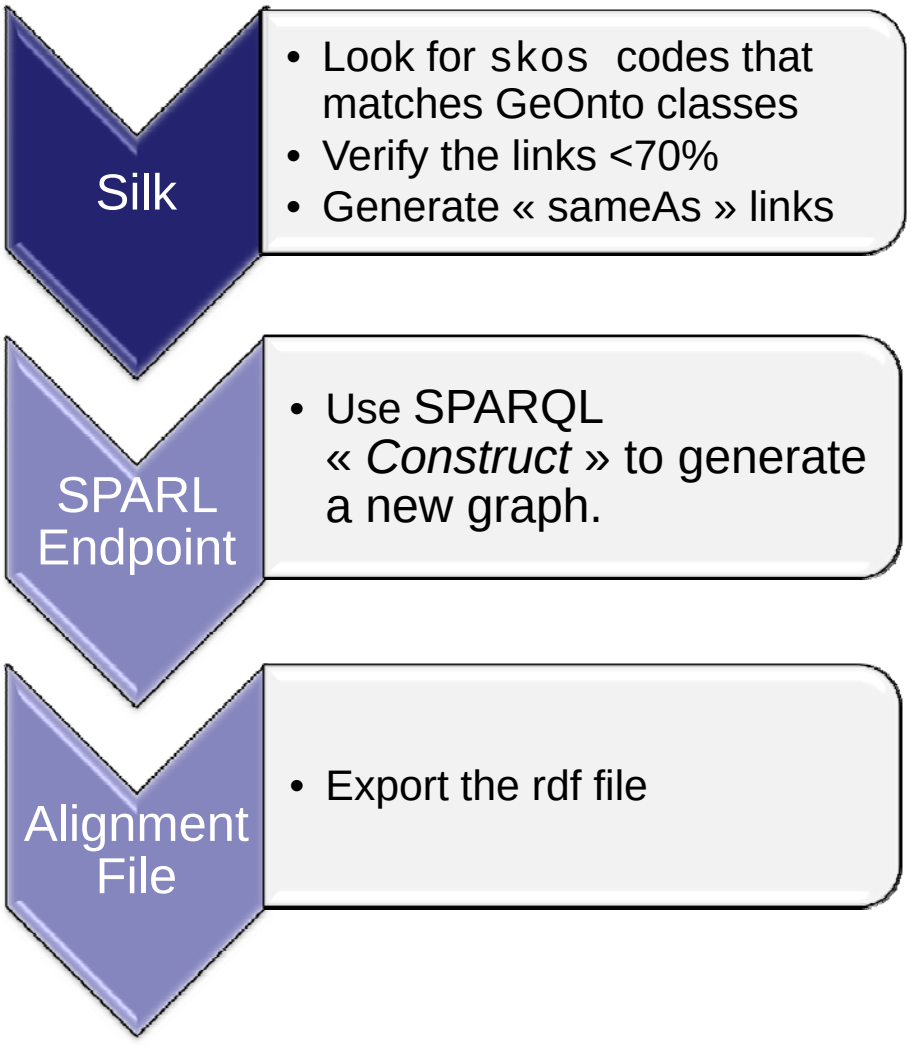


Alignment Methodology

- **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

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



Results/Evaluation

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

- **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

Topological Functions in GeoVocabs

Geo-vocabulary	Topological Functions	GeoSPARQL Requirements	Standard Followed
Ordnance Survey Spatial	`easting`, `northing`, `touches`, `within`, `contains`	Part of Req 4	OpenGIS Simple Feature
Ordnance Survey Topography	`contains`, `isContainedIn`	Very small part of Req 4	OpenGIS Simple Feature
Place Ontology	`in`, `overlaps`, `bounded_by`	Small part of Req 4	N/A
NeoGeo Spatial	All RCC8 relations	Part of Req 3; Req 6	Region Connection Calculus (RCC)
NeoGeo Geometry	—	Req 10 - 14	N/A
FAO Geopolitical	`isInGroup`, `hasBorderWith`	–	–
OntoMedia Space	`adjacent-below`, `adjacent-above`, `orbit-around`, `is_boundary-of`, `has-boundary`	–	–

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

Triple Stores and geospatial indexing

Triple store	WKT-compliance	GML-compliance	Geometry supported	Geospatial Functions	GeoVocab
Virtuoso	Yes	Yes	Point	13 functions	W3C Geo + Typed Literal
Allegro-Graph	-	–	Point	3 functions	"strip" mapping data
OWLIM-SE	–	–	Point	4 functions	W3C Geo
Open Sahara	Yes	Yes	Point, Line, Polygons	23 functions	Typed Literal
Parliament	Yes	Yes	Point, Line, Polygons	23 functions	GeoSPARQL vocabulary

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

Open Sahara
Parliament
Virtuoso

Some Recommendations

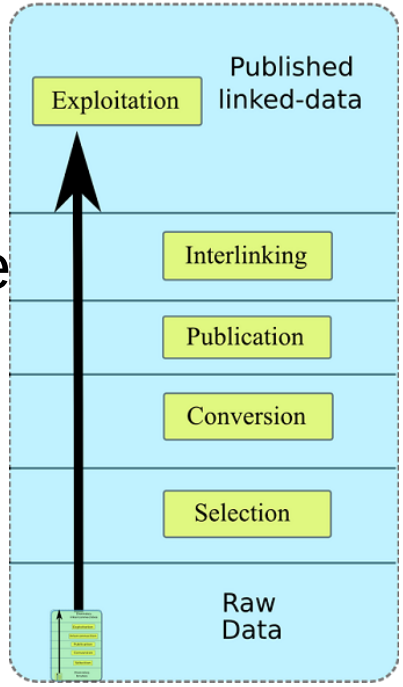
- **Complex Geometry Coverage**
  - Need to publish more data with complex geometries
  - Select suitable ontologies (e.g: NeoGeo) or GeoSPARQL
- **Features MUST be connected to Geometry**
  - Sometimes it may requires two namespaces
- **Serialization and Triple Stores**
  - Provide serialization in 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



EURECOM
Sophia Antipolis

MINES TELECOM
INSTITUT
Mines-Télécom

Thanks for your attention!

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