

Vers une meilleure interopérabilité des données géographiques françaises sur le Web de données

Ghislain Ateazing
Raphaël Troncy



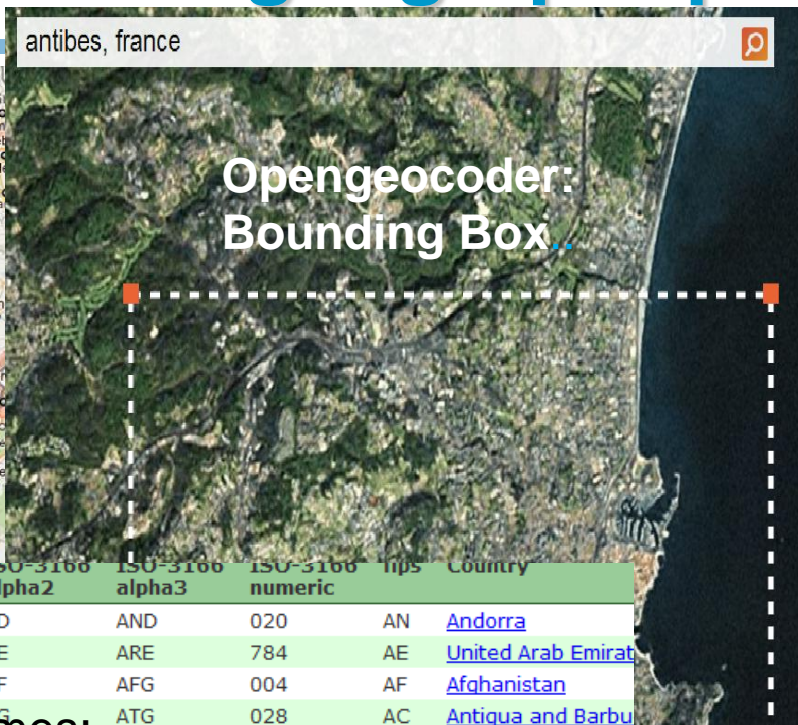
Plan

- **Où sont les données géographiques ?**
 - Le mobile et l'utilisation de la géo-localisation
 - Données géographiques sur le web sémantique
- **Modélisation des objets géographiques**
 - Modélisation des entités géographiques
 - Modélisation de la géométrie
 - Illustration avec des scénarios
- **Aligner les représentations**
- **Formuler quelques recommandations**
- **Conclusion et travaux futurs**

Où sont les données géographiques?



GeoPortail

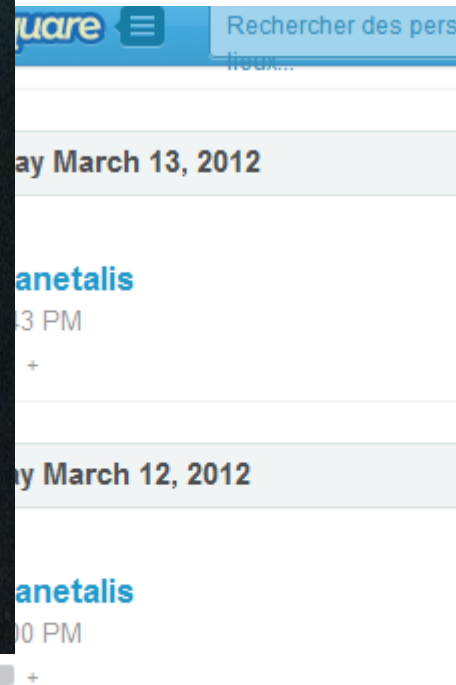


ISO-3166 alpha2	ISO-3166 alpha3	ISO-3166 numeric	tips	Country
AD	AND	020	AN	Andorra
AE	ARE	784	AE	United Arab Emirates
AF	AFG	004	AF	Afghanistan
AG	ATG	028	AC	Antigua and Barbuda
AI	AIA	660	AV	Anguilla
AL	ALB	008	AL	Albania
AM	ARM	051	AM	Armenia
AN	ANT	530	NT	
AO	AGO	024	AO	Angola
AQ	ATA	010	AY	Antarctica
AR	ARG	032	AR	Argentina
AS	ASM	016	AQ	American Samoa
AT	AUT	040	AU	Austria
AU	AUS	036	AS	Australia
AW	ABW	533	AA	Aruba
AX	ALA	248		Åland
AZ	AZE	031	AJ	Azerbaijan
BA	BIH	070	BK	Bosnia and Herzegovina
BB	BRB	052	BB	Barbados
BD	BGD	050	BG	Bangladesh
BE	BEL	056	BE	Belgium
BF	BFA	854	LIV	Burkina Faso

GeoNames:
Codes des pays

CIA Factbook

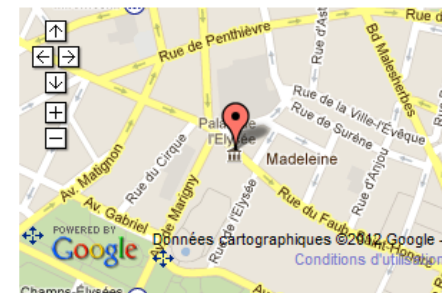
Foursquare: POI..



Adresse

Palais de l'Élysée
55, rue du Faubourg-Saint-Honoré
75008 Paris

Masquer le plan d'accès



Où sont les données géographiques liées?

- DBpedia, GeoNames



- LinkedGeodata (OpenStreetMap), Freebase (Google)
- Ordnance Survey (UK), GeoLinkedData (ES)

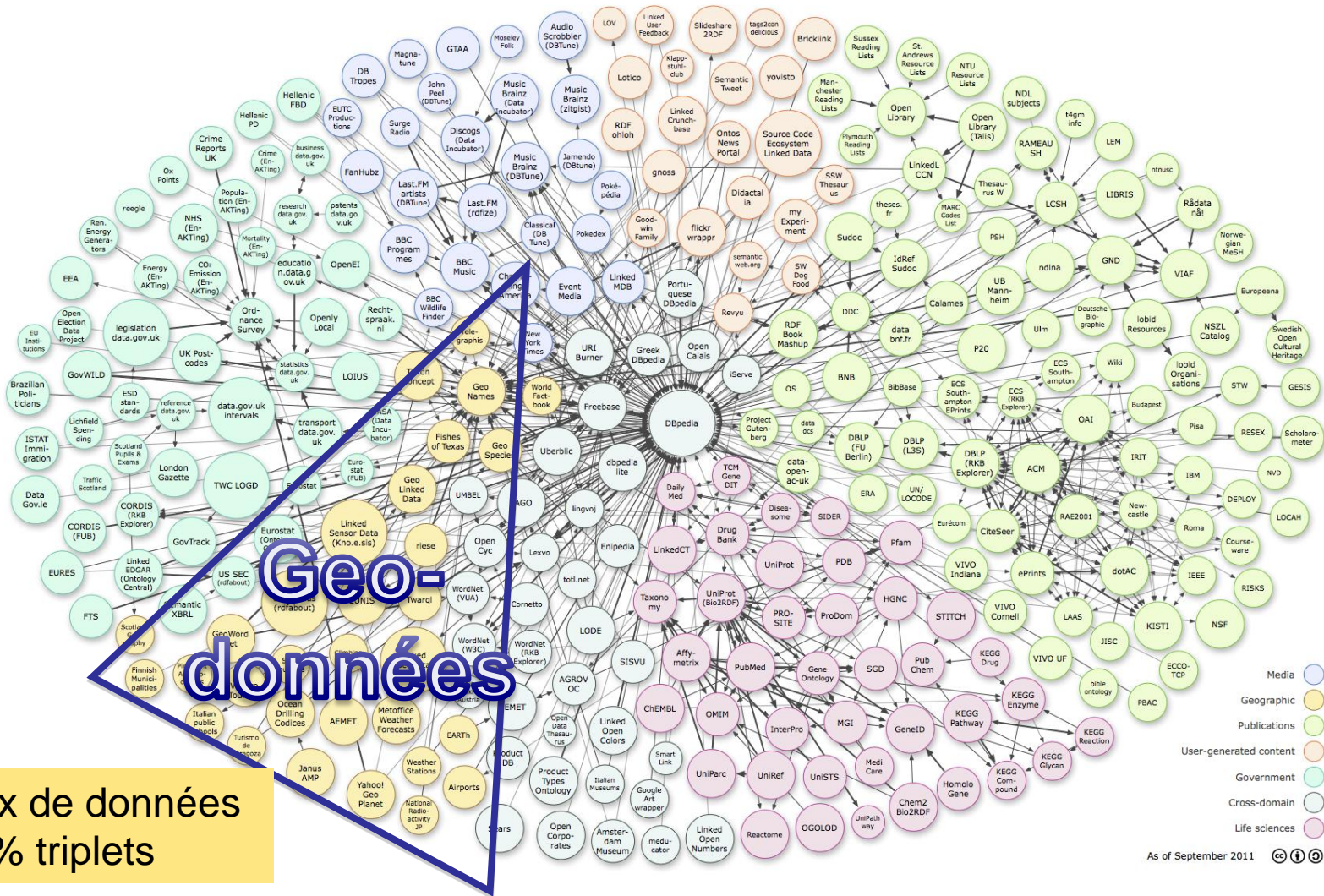
- GADM-RDF

- NUTS-RDF

- data.ign.fr (FR)

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

Données géographiques liées et LOD



31 jeux de données
 19.43% triplets

<http://lod-cloud.net/state>

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

Les vocabulaires utilisés

221 vocabularies

	Overall	Datasets
http://www.w3.org/1999/02/22-rdf-syntax-ns	126,856,808	158
http://www.w3.org/2000/01/rdf-schema	39,517,802	130
http://purl.org/dc	65,332,873	73
http://www.w3.org/2004/02/skos/core	24,546,500	60
http://www.w3.org/2002/07/owl	2,433,062	60
http://xmlns.com/foaf	6,443,666	53
http://rdfs.org/ns/void	8,180	44
http://xmlns.com/foaf/0.1	1,371,842	28
http://www.w3.org/2006/time	386,286	24
http://id.southampton.ac.uk/ns	46,355	24
http://purl.org/void	196	23
http://purl.org/dc/terms	649,031	21
http://www.w3.org/2003/01/geo/wgs84_pos	269,299	21
http://www.w3.org/2008/05/skos	8,796	7
http://purl.org/vocab	5,114	7
http://purl.org/net/provenance/ns	1,450	7
http://purl.org/linked-data/sdmx/2009/dimension	3,402,651	6
http://www.w3.org/2003/06/sw-vocab-status/ns	448	6
http://data.ordnancesurvey.co.uk/ontology/spatialrelations	267,426	5
http://www.geonames.org/ontology	29,997	5
http://purl.org/science	955,995	4

Quels sont les modèles IC utilisés ?

- **Des listes d'autorité de termes (Foursquare)**
 - Peu structurée
 - Représente les catégories des Points d'Intérêt
 - Typiquement, un type réponse d'une API
 - Nécessite d'explicitier la sémantique
- **Des catégories SKOS (GeoNames)**
 - Les classes sont des `skos:conceptScheme`
 - Les codes sont des `skos:Concept`
 - Peu de classes, de nombreux codes

Quels sont les modèles IC utilisés ?

- **Des ontologies spécifiques au domaine**
 - Les ontologies par sous-domaine (transport, unité administrative, hydrographie, etc.)
 - Ontologies interconnectées (owl:imports)
 - UK (OS) – ES (GeoLinkedData)
- **Des ontologies riches construites par des outils automatiques / TAL**
 - Les classes de haut-niveau pour structurer
 - **LinkedGeoData** (16 classes de haut-niveau; 1294 classes au total)
 - **GeOnto** (2 classes de haut-niveau; 783 classes au total)

Modélisation de la géométrie

- **Point (lat/long)**
 - Vocabulaire WGS 84 décrit au W3C
- **Rectangle (“bounding box”)**
 - Vocabulaire géopolitique (FAO)
- **Points dans une liste (« *Nodes* », « *waynodes* »)**
 - Séquence de points (LinkedGeoData)
- **Points dans un groupe de ressources RDF**
 - Un objet est “*formedBy*” ListOfPoints(GeoLinkedData.es)
- **Littéraux (représentation GML incluse en RDF)**
 - Ordnance Survey(UK)
- **Représentation complètement structurée**
 - Le vocabulaire NeoGeo (GeoVocamp)

Scénario: Tour Eiffel / Paris 7ème

Tour Eiffel



La tour Eiffel vue depuis le nord.

Caractéristiques

Localisation	Champ de Mars (Paris 7 ^e)
Coordonnées	48° 51' 30" N 2° 17' 40" E
Emprise au sol	Carré de 125 m
Hauteur :	
· Plancher du 1 ^{er} étage	57,63 m
· Plancher du 2 ^e étage	115,73 m
· Plancher du 3 ^e étage	276,13 m
· Totale avec antenne	324 m (depuis 2000)
Poids :	



La Tour Eiffel sur DBpedia (un "Building"...)

```
dbpedia:Eiffel_Tower a dbpedia-owl:Building ;  
  a <http://schema.org/Place> ; (16 "types" différent)  
  rdfs:label "Tour Eiffel"@fr ;  
  geo:lat "48.858299"^^xsd:float ; (WGS84 )  
  geo:long "2.294500"^^xsd:float ;  
  geo:geometry "POINT(2.2945 48.8583)" ;  
  
  dbpprop:buildingType "Observation tower"@en ;  
  dbpprop:elevatorCount "9"^^xsd:int ;  
  dbpprop:location dbpedia:Paris ;  
  dbpprop:isofRegion "FR-75" ;  
  dbpprop:architect dbpedia:Stephen_Sauvestre .
```

La Tour Eiffel sur Freebase (un “skyscraper”)

```
fb:en.eiffel_tower a fb:architecture.building ;  
    a fb:architecture.skyscraper ; (12 rdf:type in total)  
  
fb:architecture.skyscraper.height_with_antenna_spire_meters  
s "324.0"^^xsd:float ;  
fb:location.geocode [  
    fb:location.geocode.longitude "2.2946"^^xsd:float ;  
    fb:location.geocode.latitude "48.85839"^^xsd:float . ] ;
```

La Tour Eiffel sur GeoNames (un “S.MMT”)

```
gnr:6254976 a gn:Feature ;

  gn:name "Eiffel Tower" ;
  gn:alternateName "에펠 탑"@ko ; (en 45 langues différentes)
  gn:featureClass gn:S [
    a skos:ConceptScheme ;
    rdfs:comment "spot, building, farm, ..."@en .
  ] ;

  gn:featureCode gn:S.MMT [
    a skos:Concept ;
    rdfs:comment "a commemorative structure or statue"@en .
  ] ;

  gn:countryCode "FR" ;
  geo:lat "48.8583" ;
  geo:long "2.29452" .
```

La Tour Eiffel sur LGD (un “chemin” de 45 points)

```
lgd:way5013364 a lgdo:Building , lgdo:ManMadeTower ,  
lgdo:Attraction ;  
  rdfs:label "Wieża Eiffel'a"@pl ; (en 13 langues.)  
lgdo:layer "2";  
lgdp:building:height "301";  
lgdp:importance "international";  
lgdo:hasNodes  
<http://linkedgedata.org/triplify/way5013364/nodes>.
```

```
  rdf:_1 lgd:node33388356;  
  ..... ;  
  rdf:_10 lgd:node33388333 ;
```

(tous les 45 points du polygone)



7ème Arrondissement in DBpedia (a gml_Feature)

```
dbpedia:7th_arrondissement_of_Paris a gml:_Feature ;
    (gml n'est pas en OWL )

    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.
```

7ème 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".
```


7ème Arrondissement in LGD (a “Suburb”)

```
lgd:node248177663 a lgdo:Suburb ;  
  
rdfs:label "7th Arrondissement"@en , "7e Arrondissement" ;  
lgdo:contributor lgd:user13442 ;  
  
<http://linkedgedata.org/ontology/ref%3AINSEE> 75107 ;  
  
lgdp:alt_name "VIIe Arrondissement" ;  
  
georss:point "48.8570281 2.3201953" ;  
geo:lat 48.8570281 ;  
geo:long 2.3201953 .
```

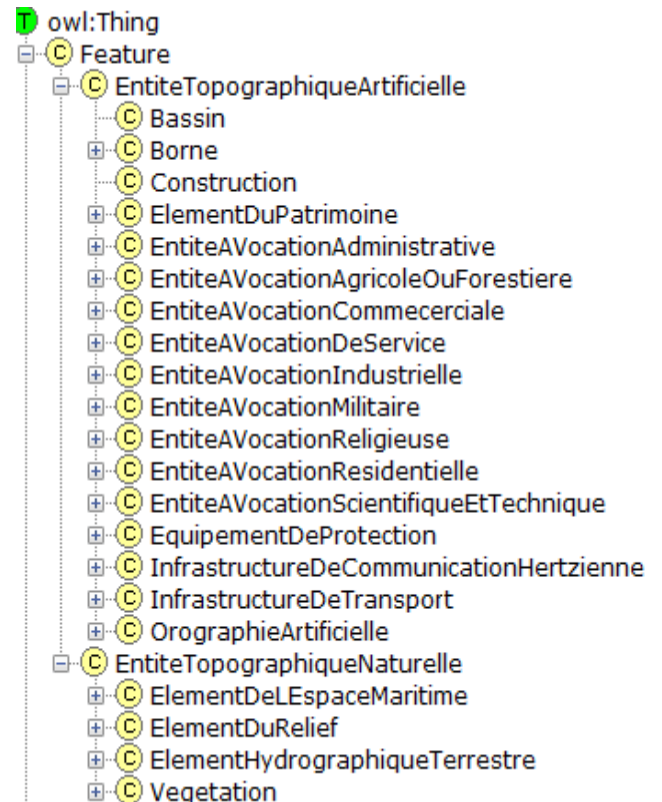
Vocabulaire GeOnto

■ Ontologie des objets géographiques (PI)

- Résultat d'un projet ANR
- Obtenue par des outils de TAL
- Recouvre la partie BDTopo (IGN)

■ Classes en français

- `rdfs:labels` en FR & EN
- Pas de `rdfs:comments`
- Peu de relations
- 783 classes

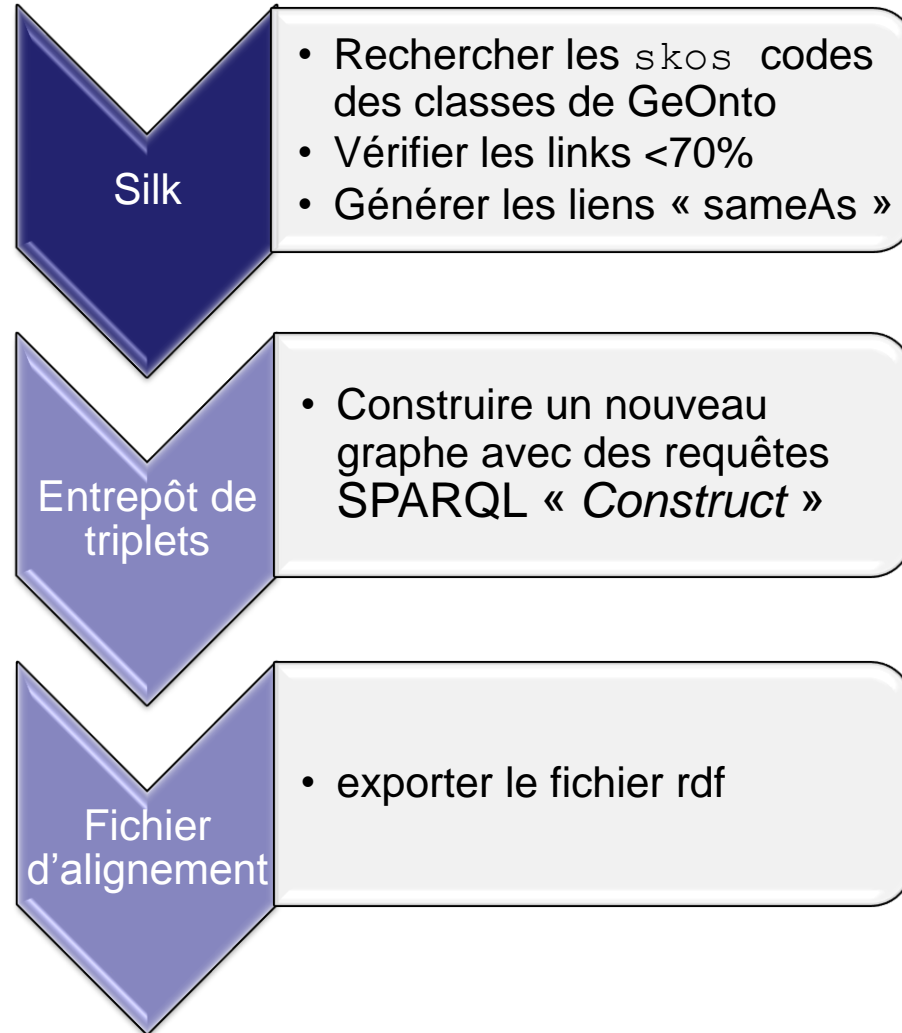


Méthodologie d'alignement

- **Alignement de GeOnto avec 4 vocabulaires et 2 taxonomies**
 - LGD, DBpedia, Schema.org, GeoNames
 - Foursquare, Google Places
- **Recherche des `owl:equivalentClass`**
 - Outil : Silk
 - Métriques : LevenshteinDistance, Jaro
 - Labels : @en des classes
 - Fonction d'agrégation : moyenne
- **Validation manuelle**
 - Pour les « `rdfs:subClassOf` »
 - Alignement spécifique avec GeoNames

Procédé d'alignement avec GeoNames

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



Résultats/Evaluation

Vocab/taxonomies	#Classes	#Classes alignées
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

- **Precisions généralement > 80%**
- **Sauf P(Schema.org) = 50%.**
 - Raison possible: les entités GeOnto plus spécifiques à la France
 - Niveau trop détaillé des entités sur Schema.org

Recommandations

- **Séparation des objets de leur géométrie**
 - Bien distinguer dans les espaces de nommage
 - Exemple: *spatial:Feature* vs *geom:Geometry*
- **Usage du vocabulaire WGS84 pour représenter les points**
 - Réutiliser les propriétés natives (long; lat; alt)
- **Géométrie complexe (polygone, courbe, etc.)**
 - Utiliser les représentations plus structurées
 - **Eviter** au maximum les nœuds anonymes dans les données
 - Proposer des transformations aux formats SIG (KML, GML, etc.)
- **Extension de GeOnto**
 - Un bon candidat: NeoGeo
 - Publier selon les Bonnes Pratiques (data.ign.fr)

Perspectives

- **Publier GeOnto++**

- Selon les Bonnes Pratiques du web de données liées

- **« Lifter » les données de l'IGN**

- En utilisant le vocabulaire

- **Poursuivre les Alignements**

- Schema.org, Foursquare, Google Place (évolutif)
- GeoSPARQL
- Procéder à des alignements au niveau des données

Merci pour votre attention !

Questions?