

GeomRDF : A Geodata Converter with a Fine-Grained Structured Representation of Geometry in the Web

Fayçal Hamdi¹ Nathalie Abadie² Bénédicte Bucher²
Abdelfettah Feliachi²

¹ISID TEAM, CEDRIC Lab
Conservatoire National des Arts et Métiers, Paris, France

²Université Paris-Est, IGN/SRIG, COGIT, Saint-Mandé, France

September 1st, 2014

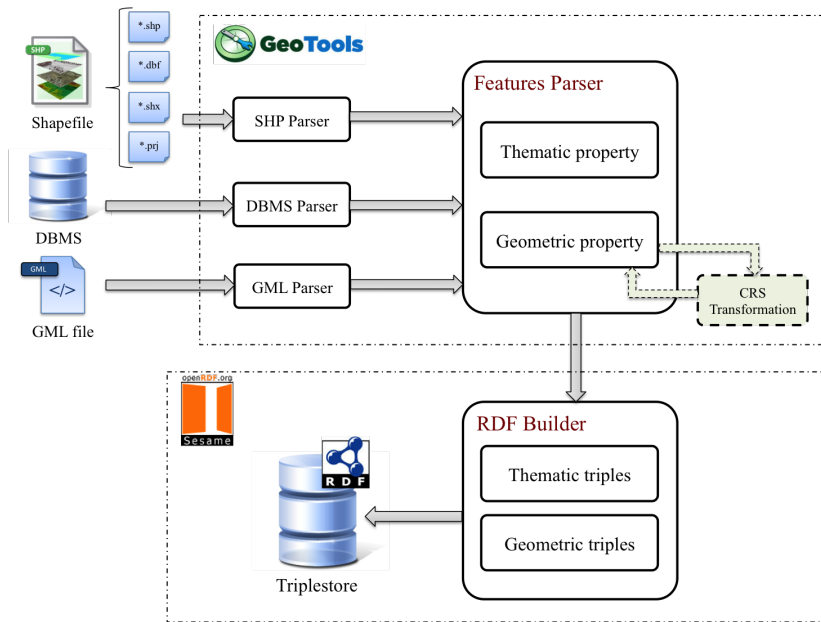
Goal

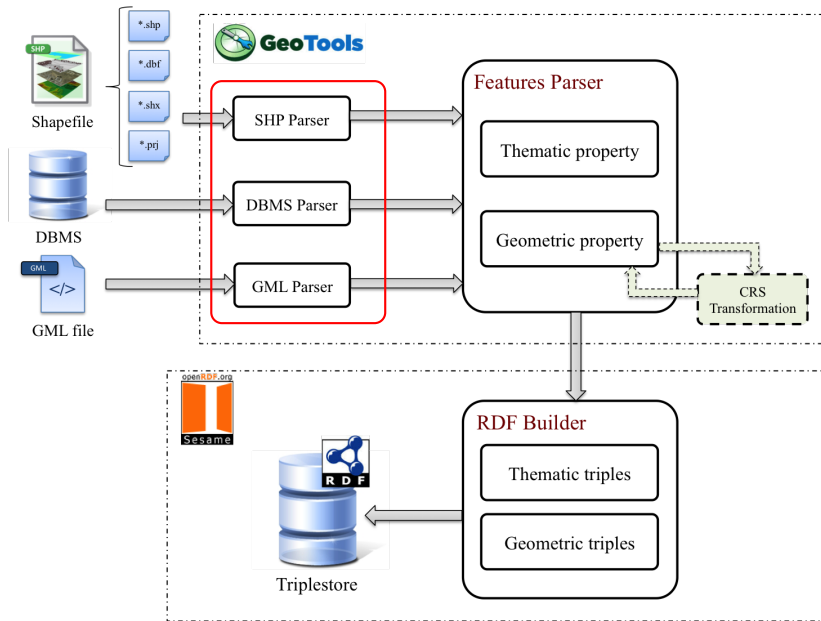
Publishers : Publishing (in an easy-way) geospatial data

Users : Querying geospatial data using merely semantic web technologies (SPARQL)

GeomRDF tool

- "Push-button" conversion requiring a minimum of user intervention
- Based on a vocabulary that reuses and extends GeoSPARQL simple feature vocabulary
- Geometries can be defined in any CRS and represented as :
 - Structured representation of geometries that can be handled with regular SPARQL queries
 - *wktLiteral* datatypes that are compliant with GeoSPARQL standard



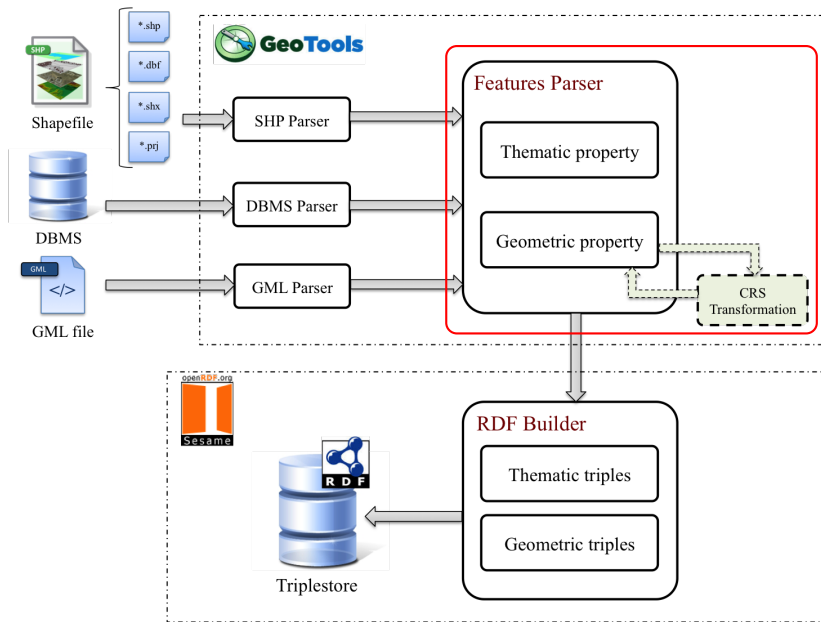


Input Parser

Definition

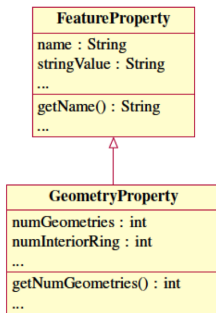
"Feature" is the "Simple Feature" defined by OGC and ISO 19125 to have both spatial and non-spatial attributes

- **ESRI Shapefile** : A spatial data format (defined by ESRI) to store, for the spatial features in a dataset, the non-topological geometry and the attribute information
- **Geospatial DBMS** : IBM DB2, H2, MySQL, Oracle Spatial, PostGIS (a spatial extension to PostgreSQL), SpatiaLite (a spatial extension to SQLite), Microsoft SQL Server and Teradata
- **GML** : OGC standard encoding specification for geodata in XML



Feature Parser

- Iterates on the extracted features and classifies their properties depending on their type
 - **Thematic Properties** : e.g., the name of the property, its value and its type
 - **Geometric Properties** : e.g., the number of geometries contained by a Multipolygon
 - Possibility of transforming coordinates from their original CRS (Coordinate Reference System) to another one



Feature parsing : the department of "Paris"

Thematic Properties :

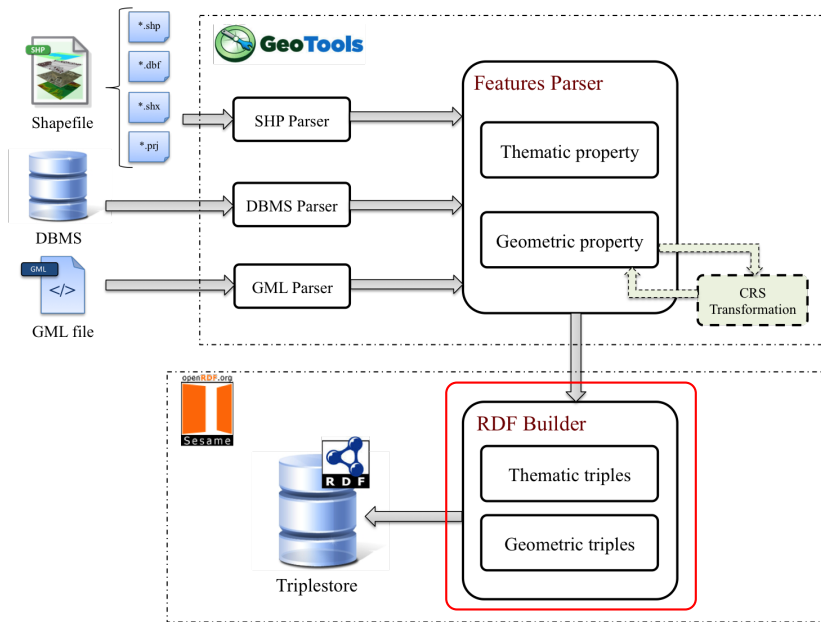
Property	CODE_DEPT	NOM_DEPT	...	CODE_REG	NOM_REGION
Value	75	PARIS	...	11	ILE DE FRANCE
Type	String	String	...	String	String

Geometric Properties :

MULTIPOLYGON(((2.41633 48.84923, 2.41597 48.84662, , 2.41633 48.84923)))

The Parser stores :

- Polygons that compose the original Multipolygon
- Exterior and the eventual Interior LinearRings that compose each Polygon
- Points included in each LinearRing
- Coordinates of each Point



RDF Builder

- Generates, from the parsed features, a collection of RDF triples
- Thematic and Geometric Properties are treated separately
 - **Thematic Properties** : The value of a property is captured as text with the appropriate RDF literal type

RDF building : the department of "Paris"

Thematic Properties :

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
@prefix geofla: <http://data.ign.fr/def/geofla#>
```

```
<http://data.ign.fr/id/geofla/departement/75> a geofla:Departement .  
<http://data.ign.fr/id/geofla/departement/75> rdfs:label "PARIS"@fr .  
<http://data.ign.fr/id/geofla/departement/75> geofla:codeDpt "75"^^xsd:string .
```

- **Geometric Properties** : The generation of the geometric RDF triples is guided by a vocabulary that extends GeoSPARQL and NeoGeo

Geometric Properties

- Triples are generated according to the Geometric Primitives Ontology : <http://data.ign.fr/def/geometrie>
- Generated RDF triples represent the geometry and the CRS of each geometric property stored by the Feature parser
- The possibility of checking the consistency and infer new axioms
- Each instance of a geometric class is linked to the URL identifying the CRS used for its coordinates

Geometric Properties (Multipolygon)

Axiom

An instance of the class `geom:MultiPolygon` is associated to one or more other instances of the class `geom:Polygon` via the property `geom:polygonMember`.

RDF building : the department of "Paris"

Geometric properties :

```
@prefix geom:<http://data.ign.fr/def/geometrie#>  
  
<http://data.ign.fr/id/geofla/departement/75> a geom:MultiPolygon ;  
geom:crs <http://data.ign.fr/id/ignf/crs/WGS84GDD> ;  
geom:polygonMember _:bn00001 ;  
geom:polygonMember _:bn00002 .
```

Geometric Properties (Polygon)

Axiom

An instance of the class `geom:Polygon` is defined by its contour which is an instance of the class `geom:LinearRing` associated with the instance of `geom:Polygon` via the property `geom:exterior`. A polygon has a single contour. It may also contain holes (0 or more), defined as instances of `geom:LinearRing` associated to `geom:Polygon` via the property `geom:interior`.

RDF building : the department of "Paris"

Geometric properties :

```
_:bn00001 a geom:Polygon ;  
geom:crs <http://data.ign.fr/id/ignf/crs/WGS84GDD> ;  
geom:exterior _:bn000021 .  
  
_:bn00002 a geom:Polygon ;  
geom:crs <http://data.ign.fr/id/ignf/crs/WGS84GDD> ;  
geom:exterior _:bn000031 .
```

The French Administrative Units (GEOFLA)

- GEOFLA : a dataset provided by the IGN France as a set of ESRI Shapefiles.
- Contains the description of all the administrative subdivisions (department, commune, arrondissement and canton) of France, overseas departments and Mayotte

The administrative subdivision	Total
Department	96
Commune	36594
Arrondissement	330
Canton	3708

The publication process

- The conversion and the publication are performed using GeomRDF and Datalift platform
 - GeomRDF is implemented as a module in the Datalift Platform (a stand-alone version of GeomRDF is also available)



Linking Datasets

Dataset Publication



Converting data formats to RDF

Vocabulary Selection

Conversion of geospatial data to RDF

- GeomRDF convert the ESRI Shapefiles to RDF
- A transformation from the original CRS of the geometries (Lambert-93) to WGS84 is carried out
- URI is constructed (according to W3C best practice) as the following :

```
<http://data.ign.fr/id/geofla/{types_of_administrative_division}/{id}>
```

Saint-Mandé commune : `<http://data.ign.fr/id/geofla/commune/94067>`

- RDF conversion based on an ontology describing the GEOFLA dataset is performed
 - Matching and replacing default predicates with the GEOFLA ontology predicates

Publishing and Interlinking

- RDF data are published online as Linked Open Data and accessible through an SPARQL endpoint address :
`http://data.ign.fr/id/sparql`
- Interlinking the administrative unit the IGN and the INSEE

Interlinking : the administrative unit of "Paris"

```
@prefix owl: <http://www.w3.org/2002/07/owl#>  
  
<http://data.ign.fr/id/geofla/arrondissement/751> owl:sameAs <http://id.insee.fr/geo/  
arrondissement/751>  
<http://data.ign.fr/id/geofla/departement/75> owl:sameAs <http://id.insee.fr/geo/departement/75>
```

Spatial queries over structured geometries

- Enable users performing basic spatial queries in SPARQL

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX geofla:<http://data.ign.fr/def/geofla#>
PREFIX geom:<http://data.ign.fr/def/geometrie#>

SELECT DISTINCT ?name WHERE {
  ?dep rdf:type geofla:Departement .
  ?dep rdfs:label ?name .
  ?dep geom:geometry/geom:polygonMember/geom:exterior/geom:points ?pl .
  ?pl rdf:type geom:PointsList .
  ?pl (rdf:rest*/rdf:first)|geom:firstAndLast ?pm .
  ?pm rdf:type geom:Point .
  ?pm geom:coordX ?x .
  ?pm geom:coordY ?y .
  FILTER ((?x > 1 && ?x < 3) && (?y > 42 && ?y < 44)) .
}
```

Conclusion

- GeomRDF a tools that enables the transformation of geospatial data into RDF
- GeomRDF supports as input different geospatial formats (ESRI Shapefile, GML and geospatial DBMS)
- The output Geometric informations are structured according to an ontology that reuses and extends the existing geographic vocabularies (GeoSPARQL and NeoGeo)
- A validation of GeomRDF has been realized against the French administrative units dataset (provided by IGN France)

Perspectives

- Extending the input format to consider text files (such as CSV) containing direct location information described by coordinates

Thank You !
Questions ?