Linking colombian open geoinformation

Diana Rocío Galindo González

July 10, 2012

1 Introduction

Several data sets of geoinformation of the countries are available on the Internet. From local studies to national or regional agencies dispose vector data, satellite images and maps in different formats. The increasing of volunteer geographic information also enlarge the access of this type of information in kml files, gps tracks among other formats.

Most of the sources of information are dispersed, in the particular case of colombian geoinformation the access to this information is still indirect, resources are not available in explicit links, queries can be complex according to each institution and most of the studies are constrained for lack of datasets or geographic information(Figure 1).

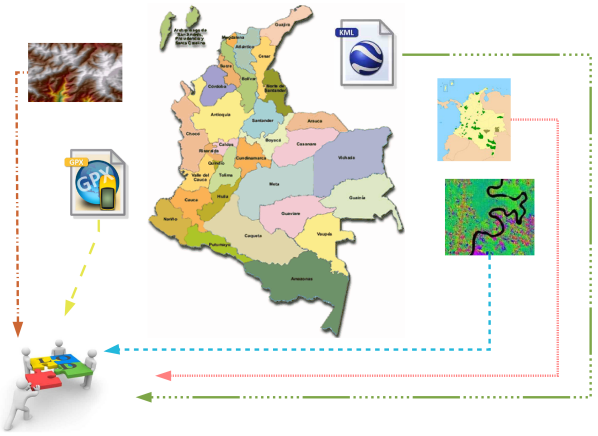

Figure 1: Objective of the project

Linked open data provide the possibility to relate and dispose information existing on the web through Resource Description Framework (RDF). Geographic information has been also modelled using this format, a clear example is GeoLinkedData [Auer et al., 2009] based on OpenStreetMap structure and information to provide a browser of linked information in RDF format. Unlike GeoLinkedData, this project does not aim to include a map viewer, but to provide users a list with different links to download open geographic information from

Colombia. As a result,this project will simplify the access of such information through queries by specific administrative entities of Colombia.

Colombia is a country located in South America, its political and administrative division contains 33 departments and 1121 municipalities. The country is object of study in several fields due to its environmental, biological and cultural diversity. Geoinformation plays both a basic or complementary role to perform those analysis. Most of the research make by students have constraints to obtain data sets, geographic information is expensive or not easily accessible.

The project aims to link different sources of colombian open geoinformation through the disposal of links accesible via SPARQL queries of departments or municipalities. This is a simple application of linked open data to collect and display information from different sources applied for a specific country, but could be applied to other regions or using different divisions.

2 Sources of geoinformation

There was two main sources of geoinformation for this project:

- ◇ *From Global Land Cover Facility*[1]:This service provides along several products remotely sensed satellite data.
- ◇ *From The University of Texas - Perry-Castañeda Library Map Collection*[2]:This website dispose maps from different sources and formats for various countries including Colombia.

Each of the data sets were characterized and linked using Resource Description Framework (RDF).

3 Classes and properties

The longest part of the exercise was the modelling of the relation between satellite imagery available in Global Land Cover Facility (GLCF) and different regions in Colombia. It was necessary to convert the online database to RDF format and link it with the administrative boundaries. In this case, the images and the departments were the classes, only the information of departments of Colombia was linked on DBpedia. Each type of the datasets also were considered classes.

In general, properties as administrative boundaries have been considered in ontologies [Kauppinen, 2010] and translated into vocabularies allowing to use them in the project. For this case, were collected some terms of different vocabularies to characterize each dataset of geoinformation encountered on the internet.

Some properties available in different vocabularies used in this project were:

[1] http://glcf.umiacs.umd.edu/
[2] http://www.lib.utexas.edu/maps/colombia.html

- ◇ *From DCMI Metadata Terms*[3]: To describe the date of the datasets *date*, its identifier *issued* and to summarize the information about the dataset *description*.
- ◇ *From NeoGeo Vocabulary*[4]: To describe the geometry in case of shapefile format *geometry*, and to describe relation between datasets *overlaps*.
- ◇ *From Open Time and Space Core Vocabulary Specification*[5]: To describe the relation between the geoinformation datasets and, in general, between the classes, space properties: *northOf, northWestOf, westOf, southWestOf, southOf, southEastOf, eastOf* and *northEastOf*. To characterize the datasets and the sensor in case of remote sensing satellite images, temporal properties: *happensAt, existenceBeginsAt, existenceEndsAt* and *inFutureOf*.
- ◇ *From Friend of a Friend Vocabulary Specification*[6]: To describe the type of the data *Document*

The figure 2 shows an example of a triple using the existent vocabulary.

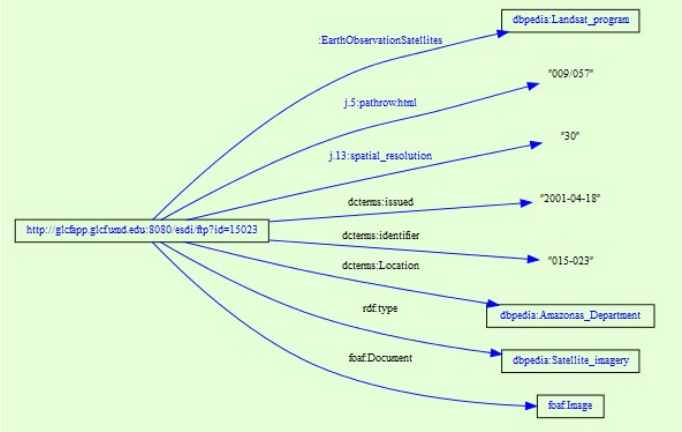

Figure 2: Example of triple using existing vocabularies

Vocabulary to characterize properties from satellite images are still insufficient, some of the properties were defined linking some resources on the internet.

4 Results

As was mentioned before the primary data set was the GLCF images, in the process, 1.421 satellite images were found covering Colombia in different dates. For this work only were used 50 images as example, each satellite image was characterized with at least 10 properties (triples). The same procedure was

[3]http://purl.org/dc/terms/
[4]http://geovocab.org/doc/neogeo/
[5]http://observedchange.com/tisc/ns/
[6]http://xmlns.com/foaf/spec/

used to other geoinformation sources, but this kind of information required less properties to be characterized.

Here, some examples of the generated triples:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dcterms:      <http://purl.org/dc/terms/> .
@prefix dbpedia:  <http://dbpedia.org/resource/> .
@prefix foaf: <http://sws.geonames.org/> .
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:     <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dcat:     <http://www.w3.org/ns/dcat#> .

#Classes

#Class Satellite image

<http://dbpedia.org/page/Satellite_imagery> <rdf:type> <owl:Class>.
<http://dbpedia.org/page/Satellite_imagery> <rdf:label> "Satellite Image".

#Class Department

<http://dbpedia.org/page/Department_%28country_subdivision%29>
<rdf:type><owl:Class>.
<http://dbpedia.org/page/Department_%28country_subdivision%29>
<rdf:label> "Department".
```

for a satellite image:

```
#Properties

#Property path_row

<http://glcf.umd.edu/esdi2-help/pathrow.html>
<rdf:type><dbpediaowl:DatatypeProperty>.
<http://glcf.umd.edu/esdi2-help/pathrow.html>
<rdfs:range><xsd:string>.
<http://glcf.umd.edu/esdi2-help/pathrow.html>
<rdfs:label>"Path/Row"^^<xsd:string>.

#Property spatial resolution
<http://dbpedia.org/page/spatial_resolution>
<rdf:type> <dbpediaowl:DatatypeProperty>.
<http://dbpedia.org/page/spatial_resolution>
<rdfs:range> <xsd:int>.
<http://dbpedia.org/page/spatial_resolution>
<rdfs:label>"Spatial resolution in meters"^^<xsd:string>.

<dbpedia:Satellite_imagery>
<dcat:accessURL><http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021> .
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<foaf:Document><foaf:Image>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<rdf:type> <dbpedia:Satellite_imagery>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<dcterms:identifier> "015-021 "^^<xsd:ID>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<http://glcf.umd.edu/esdi2-help/pathrow.html>"008/058"^^<xsd:string>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<http://dbpedia.org/page/spatial_resolution>"30"^^<xsd:int>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<dcterms:issued> "2002-01-08"^^<xsd:date>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<dcterms:Location> <dbpedia:Caqueta_Department>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<dcterms:Location> <dbpedia:Huila_Department>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<dcterms:Location> <dbpedia:Meta_Department>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<dcterms:Location> <dbpedia:Tolima_Department>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<dbpedia:class/yago/EarthObservationSatellites>
```

```
<dbpedia:Landsat_program>.
<http://glcfapp.glcf.umd.edu:8080/esdi/ftp?id=15021>
<dc:description> "Satellite image of sensor ETM+ with30 of spatial resolution"^^<xsd:string>.
```

for a kml file:

```
<http://www.peakware.com/xml/map.php?type=kml&map=area&id=298>
<rdf:type> <dbpedia:Keyhole_Markup_Language>.
<http://www.peakware.com/xml/map.php?type=kml&map=area&id=298>
<dcterms:Location> <dbpedia:Magdalena_Department>.
<http://www.peakware.com/xml/map.php?type=kml&map=area&id=298>
<dcterms:Location> <dbpedia:Tolima_Department>.
<http://www.peakware.com/xml/map.php?type=kml&map=area&id=298>
<dcterms:Location> <dbpedia:Risaralda_Department>.
<http://www.peakware.com/xml/map.php?type=kml&map=area&id=298>
<dcterms:Location> <dbpedia:Quindio_Department>.
<http://www.peakware.com/xml/map.php?type=kml&map=area&id=298>
<dcterms:Location> <Valle_del_Cauca_Department>.
<http://www.peakware.com/xml/map.php?type=kml&map=area&id=298>
<dc:description> "KML of peaks over colombian andes"^^<xsd:string>.
```

The complete set of generated triples for this project can be observed at the student etherpad [7] used to store the information and produce RDF serialization through the service of University of Münster RDFpad from LODUM group [8].

5 SPARQL query

The second part of the exercise consisted in perform a query of the resources according to its location, in this case, the department level as a easy way to locate an area of interest.Once the information was structured in the RDF format, it was possible to access to it through a SPARQL query, an example could be seen in the Figure 3. The query was done using SPARQLfly endpoint also from LODUM at University of Muenster.

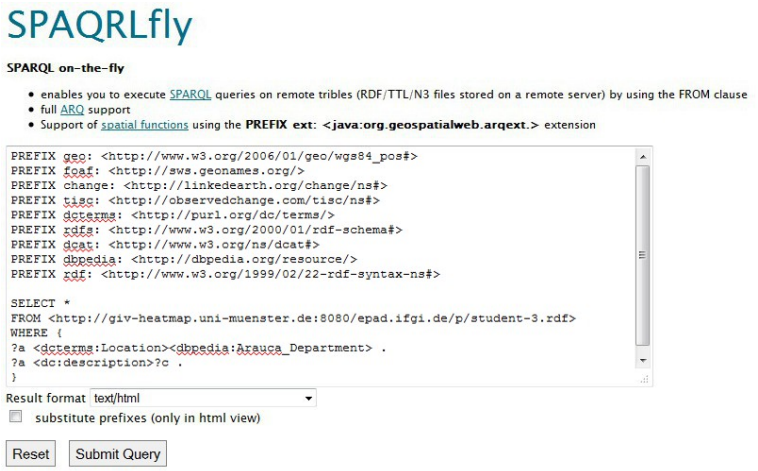

Figure 3: Example of SPARQL query.

[7] http://epad.ifgi.de/p/student-3
[8] http://http://lodum.de/

The difference with a query over a single data sets available on internet, is that the user is including different resources at the same time instead one single data set, for example the satellite imagery, just using an easy identifiable name as is the department. As a result, the query shows a small description of the data set and the link where the resource can be downloaded given a department of interest (See Figure 4).

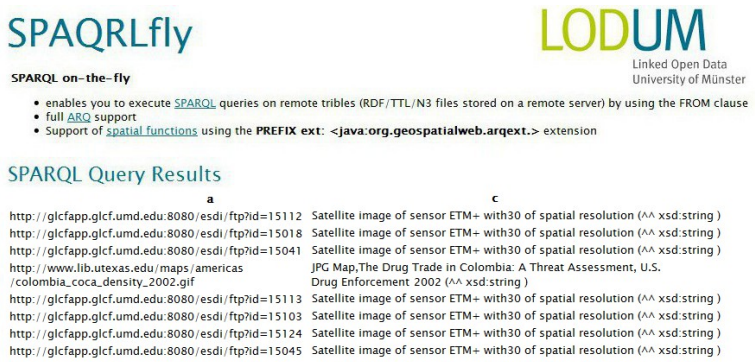

Figure 4: Result of SPARQL query.

6 Conclusions

From the process were obtained several conclusions, the most important is that open geoinformation available on the Internet can be easily connected and disposed through linked open data, also can model using an easier approach of the location of each data set compared with the offered by the websites of the original data sets. In this case the name of the department was linked to the resource and the user could access faster to several datasets at the same time instead to do an individual query per resource.

The colombian case was appropriate to show the benefits of dispose geoinformation in RDF format. The information of the departments of the country is available on dbpedia and also bring the possibility to obtain further information of each region.

The existing vocabulary allows spatial and temporal characterization of the data, but still is required a specific vocabulary for geoinformation in raster model as the satellite imagery and its associated characteristics.

References

[Auer et al., 2009] Auer, S., Lehmann, J., and Hellmann, S. (2009). Linkedgeodata: Adding a spatial dimension to the web of data. *The Semantic Web-ISWC 2009*, pages 731–746.

[Kauppinen, 2010] Kauppinen, T. (2010). *Methods for Creating and Using Geospatio-temporal Semantic Web*. PhD thesis, Aalto University, School of Science and Technology, Espoo.