PlanetData

Network of Excellence

FP7 – 257641

D16.3 Call 2: Linked Map Authoritative dataset

Coordinator: Jesús Barrera (GEOSLAB)
With contributions from: Francisco J Lopez-Pellicer (UNIZAR)

1st Quality reviewer: Giorgos Flouris (FORTH)
2nd Quality reviewer: Aljaz Kosmerlj (JSI)

Deliverable nature:	Other (O)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	M41
Actual delivery date:	M43
Version:	1.0
Total number of pages:	34
Keywords:	Linked data, geographic information, maps, dataset

Abstract

This deliverable describes the first version of the production as Linked Data of the BCN25/BTN25 dataset published by the Spanish National Geographic Institute (IGN), and its enrichment with links to resources published as part of Volunteer Geographic Information datasets.

Executive summary

The WP 16 is focused, among other objectives, on the conversion of the governmental geographic dataset BCN25/BTN25 published by the Spanish National Geographic Institute (IGN) into Linked Data, the development of a provisioning mechanism for Volunteer Geographic Information (VGI) datasets, the development of a schema for storing provenance, and the creation and maintenance of links between BCN25/BTN25 and VGI datasets and their updates. Part of these objectives were fulfilled with D16.1 "*Call 2: Linked Map VGI provenance schema*" and D16.2 "*Call 2: Linked Map Provisioning service*". D16.1 describes a schema based on the PROV Data Model built for describing the provenance of each dataset processed in the Linked Map project. D16.2 describes the prototype of a system created with the aim of provisioning and keeping synchronized a local RDF storage with geographic datasets from governmental and Volunteer Geographic Information (VGI) sources.

This deliverable presents the results obtained after applying the processes and schemas described in D16.1 and D16.2 for the production as Linked Data of the BCN25/BTN25 dataset and, in addition, describes the linking system that is responsible for discovering relationships between the processed data and setting RDF links between data belonging to different VGI sources.

First dumps of the dataset will be available at Linked Map web site[1], along with the source code used in their production[2].

[1] http://linkedmap.unizar.es/datasets/
[2] http://linkedmap.unizar.es/technology/

Document Information

IST Project Number	FP7 – 257641	**Acronym**	PlanetData
Full Title	PlanetData		
Project URL	http://www.planet-data.eu/		
Document URL	http://wiki.planet-data.eu/web/D16.3		
EU Project Officer	Leonhard Maqua		

Deliverable	**Number**	D16.3	**Title**	Call2: Linked Map Authoritative dataset
Work Package	**Number**	WP16	**Title**	Call2: Linked Map Linked data provisioning

Date of Delivery	**Contractual**	M41	**Actual**	M43
Status	version 1.0		final ☒	
Nature	prototype □ report □ demonstrator □ other ☒			
Dissemination level	public ☒ restricted to group □ restricted to programme □ consortium □			

Authors (Partner)	Jesús Barrera (GEOSLAB), Francisco J Lopez-Pellicer (UNIZAR)			
Responsible Author	**Name**	Jesús Barrera	**E-mail**	jesusb@geoslab.com
	Partner	GEOSLAB	**Phone**	+34 976 065152

Abstract (for dissemination)	This deliverable describes the first version of the production as Linked Data of the BCN25/BTN25 dataset published by the Spanish National Geographic Institute (IGN), and its enrichment with links to resources published as part of Volunteer Geographic Information datasets.
Keywords	Linked data, geographic information, maps, dataset

Version Log			
Issue Date	**Rev. No.**	**Author**	**Change**
2014/03/10	0.1	Jesús Barrera Francés	Template instantiation
2014/03/26	0.4	Jesús Barrera Francés	First version
2014/04/03	0.5	Francisco J Lopez-Pellicer	Version for QA
2014/04/17	0.6	Francisco J Lopez-Pellicer	Version after FORTH review
2014/04/24	0.9	Francisco J Lopez-Pellicer	Version after JSI review
2014/04/25	1.0	Jesús Barrera Francés	Final version

Table of Contents

Executive summary 3
Document Information 4
Table of Contents 5
Abbreviations 6
List of figures 7
List of tables 8
1 Introduction 9
2 Provisioning service results 10
2.1 Namespaces 10
2.2 From raw files to a spatial database 11
2.3 From a spatial database to a RDF store 13
3 Linking system 15
3.1 Namespaces 15
3.2 Silk configuration 15
3.3 Pairing tasks 15
4 Conclusions 18
Appendix I BCN25/BTN25 feature type selection 19
Appendix II Example of PROV-XML template 21
Appendix III Example of R2RML File 23
Appendix IV Silk project Example 26
Appendix V Geokettle configuration file 28
References 34

Abbreviations

BCN25 Base Cartográfica Numérica 1:25.000 (Numeric Cartographic Base 1:25.000)

BTN25 Base Topográfica Nacional 1:25.000 (National Topographic Base 1:25.000)

GI Geographic Information

OGC Open Geospatial Consortium

SDB Spatial Database

RDB Relational Database

VGI Volunteer Geographic Information

List of figures

Figure 1 – Service Architecture 9
Figure 2 – Provisioning system architecture 10
Figure 3 – Database model 11
Figure 4 – SHP2RDB Transformation 12
Figure 5 – MDB2RDB Transformation 13
Figure 6 – RDF model 13
Figure 7 – BCN25 Tile Partitions 16
Figure 8 – BCN25-DBPedia comparison task 17

List of tables

Table 1 – BCN25/BTN25 feature types selection 19

1 Introduction

The Linked Map project aims to demonstrate the enormous potential of linking together different sources of geographic information under the paradigm of Linked Data. To achieve its objectives, the project will carry out the conversion of one or several governmental datasets to Linked Data so it can be linked to VGI sources (OpenStreetMap, DBpedia, etc.). The selected official or governmental datasets (that in the context of the project are called authoritative dataset) to be processed in the project are the Numerical Cartographic Database (BCN25/BTN25) and its derived product the *Nomenclátor Geográfico Básico de España* (NGBE). WP 16 is the responsible of carrying out the task: the provisioning of an authoritative geographic dataset and several VGI sources, their transformation into RDF data and their integration using RDF links.

Figure 1 shows an overall view of the systems involved in the tasks defined in WP 16. The first one is the provisioning system, which was described in the deliverable D16.2 "*Call 2: Linked Map Provisioning Service*" [1]. It makes the conversion from the original sources to RDF that will be stored in the data backend. This repository is used by the linking system, which is the responsible of discovering and establishing the links between the features from different datasets. Both elements, the final results stored in the data backend, which have been obtained after applying the processes of the provisioning service, and the linking system, are the topics described in the current document. Finally, the Linked Map service will offer a GeoSPARQL node that will provide access/update to the information stored in the data backend. This service will be described in the deliverable D16.4 "*Call 2: Linked Map Data access/update service*" [2].

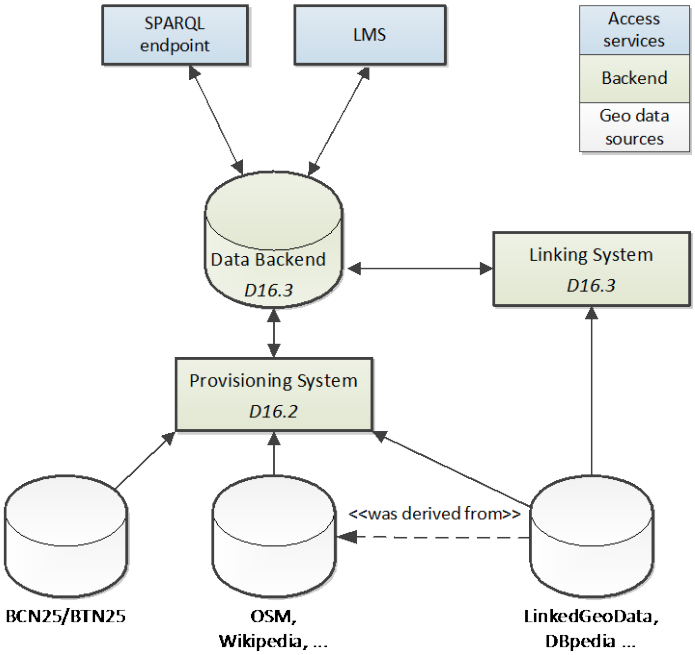

Figure 1 – Service Architecture

This document describes how an authoritative geographic dataset (BCN25/BTN25) has been transformed into RDF. This process was outlined in the deliverable D15.1 "*Call 2: Linked Map requirements definition and conceptual architecture*" [3]. The document is organized into two main sections. Section 2 collects the results obtained after putting into practice the provisioning service described in the deliverable D16.2 through the two main processes that compound this service, that is to say, the process that transforms the source datasets into a common model in a relational database and the process that converts the information stored in this database into RDF. Section 3 describes the linking system, which has been developed using Silk [4], an open source solution for discovering relationships between data from different sources.

2 Provisioning service results

Figure 2 shows the main architecture of the provisioning service already described in D16.2. Basically it distinguishes three main steps conforming this service. The first step has been carried out by external providers (IGN/CNIG[3], OSM Foundation[4], Geofabrik[5]…). They have transformed and published their datasets in Esri shapefile [5] and Microsoft Access formats. In order to register data provenance, we only need to produce some associated documentation. The second step extracts data from a spatial datasets stored in Esri shapefile and Microsoft Access formats and loads them into a relational model in a spatial database. Finally the last step extracts data from this spatial database, transforms them into RDF format, and loads them into a spatial RDF store that supports the GeoSPARQL vocabulary. Besides these three processes, there is another additional one that is part of all of them. It manages and updates the provenance information about the datasets by registering the transformations made on the data using the provenance model described in D16.1 "*Call 2: Linked Map VGI provenance schema*" [6].

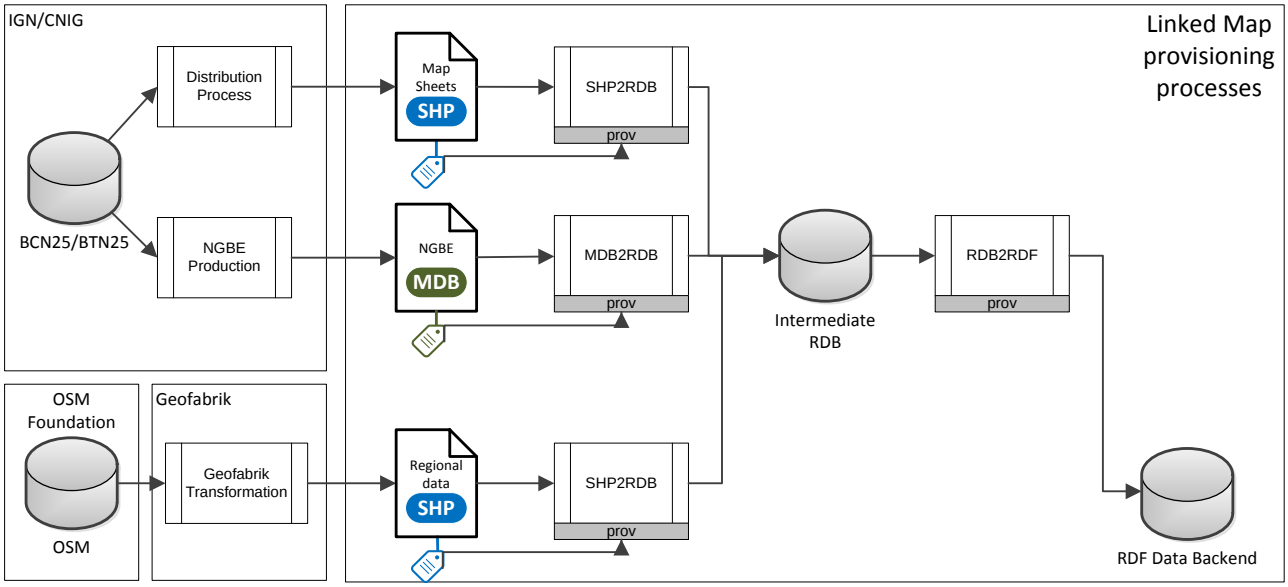

Figure 2 – Provisioning system architecture

The above processes had been applied to the official datasets (BCN25/BTN25 and NGBE) and the OpenStreetMap dataset. Also Wikipedia data have been taken into account in the project, but they have been directly linked through the DBpedia SPARQL live endpoint as showed in Section 3.

In D16.2 was discussed why we follow this two-step approach instead of a direct transformation from spatial data formats to RDF. The root cause is that the analysed spatial to RDF tools do not completely fulfill the requirements of the Linked Map project.

The following sections provide details of the results obtained after applying the above processes.

2.1 Namespaces

The following namespaces and prefixes will be used in this section:

```
geo:   http://www.opengis.net/ont/geosparql#
rdf:   http://www.w3.org/2000/01/rdf-schema#
rdfs:  http://www.w3.org/1999/02/22-rdf-syntax-ns#
dct:   http://purl.org/dc/terms/
prov:  http://www.w3.org/ns/prov#
```

[3] http://www.ign.es

[4] http://www.openstreetmap.org/

[5] http://www.geofabrik.de/

```
li:    http://linkedmap.unizar.es/def/prov-li#
ty:    http://linkedmap.unizar.es/id/types/
dt:    http://linkedmap.unizar.es/id/data/
pr:    http://linkedmap.unizar.es/id/provs/
ds:    http://linkedmap.unizar.es/id/datasets/
```

2.2 From raw files to a spatial database

Figure 3 shows the database model design. It consists of a common table storing metadata about all processed datasets and an individual table to store the data of each dataset. DATASETS table stores metadata about the name of the dataset and information about provenance. Individual tables have a set of common attributes among them (feature identifier, feature type, feature label, geometry and provenance) and some additional attributes, which are specific in the context of the dataset itself. It is worth noting that a dataset can be related with several individual tables. For instance, the BCN25/BTN25 needs a different table for each feature type because they can have a different data model. Relation between the DATASETS table and the tables storing the data is made through the table named DATASETS_TABLES_REL.

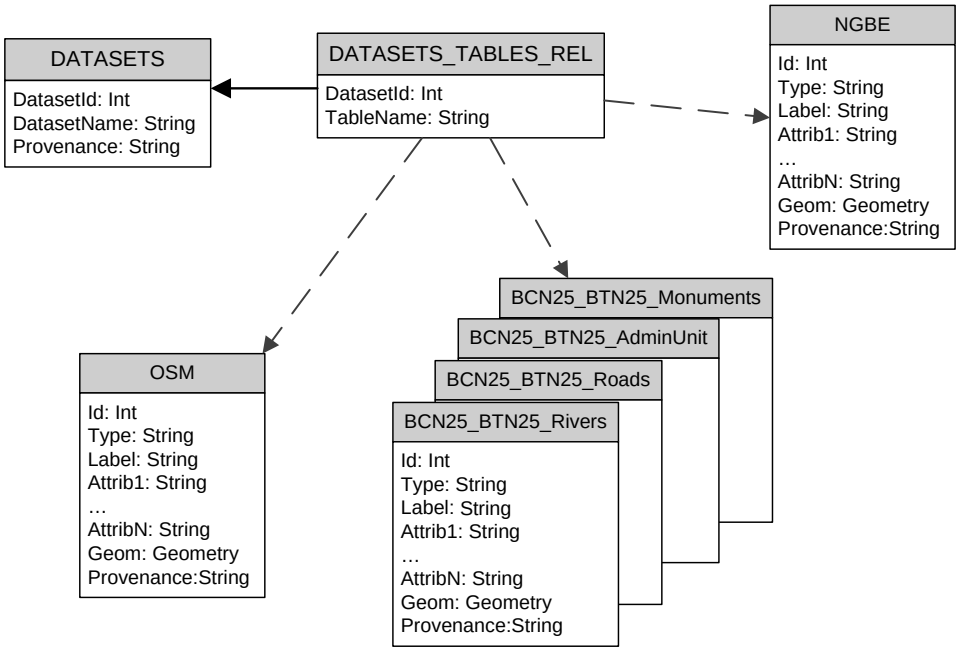

Figure 3 – Database model

As seen in the deliverable D16.2, the selected tool to make this transformation is GeoKettle [7], a powerful spatial Extract, Transform and Load (ETL) solution that can work with a wide range of spatial data formats, performing several transformations and converting them into many other different formats. Every GeoKettle Transformation is created with a graphical tool so it is fair to describe GeoKettle as metadata oriented data-flow tool.

Two different kind of GeoKettle Transformations have been created:

- SHP2RDB: This kind of transformation extracts geographic information from Esri shapefiles, transforms it to a common model and stores it into a database. The transformation is parameterized and acts as a template that is instantiated for transforming different Esri shapefiles with different thematic content.
- MDB2RDB: This transformation extracts geographic information from Microsoft Access files, transforms it to a common model and stores it into a database.

Both transformations load the results into a PostgreSQL/PostGIS [8] database including information about the data provenance.

The SHP2RDB Transformation (Figure 4) takes as input an Esri shapefile which is composed of at least three individual files: a main file (`.shp`) that contains the spatial geometries, a dBASE table file (`.dbf`) with a row for each feature in the dataset containing its alphanumeric attributes and an index file (`.shx`) with indexing information. Only the first two of these files are used in this process. The dBASE file is processed in order to extract the `ID`, `TYPE` and `LABEL` common attributes as well as all the specific ones depending on the source dataset. The main file is used to obtain the geometries that next are converted into PostGIS geometries. Finally, provenance information describing from where the information has been extracted is added encoded in PROV-XML format [9]. The GeoKettle Transformations use a set of PROV-XML templates for adding to each feature a description of its provenance. Templates include variable parts that are instantiated in runtime. Appendix II presents one of the provenance templates used.

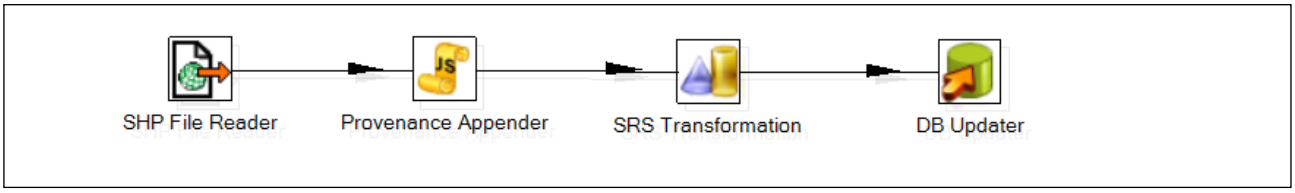

Figure 4 – SHP2RDB Transformation

The SHP2RDB Transformation has been used for processing two different datasets: The BCN25/BTN25 and OpenStreeetMap.

The *Base Cartográfica Numérica* scale 1:25000 (Numerical Cartographic Database, BCN25) [10] and the *Base Topográfica Nacional* scale 1:25000 (National Topographic Database, BTN25) [11] are two geographic databases developed by the National Geographic Institute of Spain (IGN) that provides a continuous coverage of Spain. The second one is a replacement for the first one and both datasets are available for download at the *Centro de Descargas* in Esri shapefile format.

The BCN25/BTN25 dataset is divided into map sheets that contain one shapefile for each feature type. Due to the fact that these shapefiles can have different data models, the SHP2RDB GeoKettle process has been run separately for each kind of shapefile and each one of them has been considered as a new dataset. The result is that every feature type has been modelled in a separate table in the database and has a specific row in the `DATASETS` table. As a result of the execution, 4042 ZIP files (about 9 GB of compressed information) have been processed using 50 different GeoKettle Transformations (see an example in Appendix V). This process has inserted about 800.000 features into the database.

Not all feature types from the BCN25/BTN25 have been processed (see rationale in deliverable D16.2). We have established some criteria (or rules) for selecting the more interesting features in the context of the project and these criteria have been followed when applying the provisioning procedures. The main rule is that feature types that contain a name should be processed in the first place, due to the fact that we are going to link the features in the BCN25/BTN25 with features in Wikipedia. The second rule is that feature types used in OpenStreetMap must be considered in second place, because we want to link both datasets. Thus, features without a name and belonging to a feature type that is not considered in OpenStreetMap have not been taken into account in the context of Linked Map. The result of this selection is presented in Appendix I, which lists all feature types contained in the BCN25/BTN25 noting if they are selected or not to be processed in the project.

The last dataset processed in the project has been OpenStreetMap (OSM). In relation to the provisioning of OSM datasets, the LinkedGeoData[6] project converted OSM data to RDF and provided a light ontology from it. Although the data can be downloaded, accessed via a REST interface or as Linked Data, and queried through a SPARQL endpoint, this dataset has not been updated for a year and the live service offered to access to OSM data is down. As consequence, we discard LinkedGeoData as a source of OSM data. After a careful analysis, we preferred to use snapshots extracted daily from the OSM database and published in Esri shapefile format by Geofabrik as source of OSM data. This approach also enables the use of the SHP2RDB GeoKettle Transformation. This process has extracted about 350.000 features from OSM belonging to 8 different feature types.

[6] http://linkedgeodata.org/

The MDB2RDB Transformation (Figure 5) is composed by 6 steps: the first one extracts table information that is stored as text in Microsoft Access, the second one generates a PostGIS geometry from pairs (X,Y) of coordinates stored in the original file, the third step will add the generated geometry to the rest of the properties. In the fourth step, the GeoKettle Transformation assigns a reference system to the generated geometry. In the fifth phase of the process provenance information is appended to the rest of properties and finally the sixth step stores the information into the database. Finally, just as the SHR2RDB process does, provenance information is added via the instantiation of a PROV-XML template.

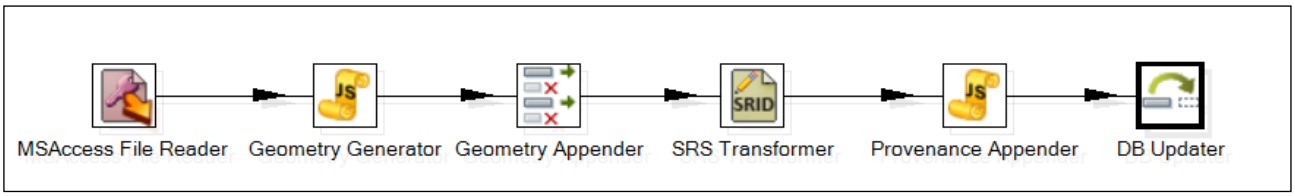

Figure 5 – MDB2RDB Transformation

The MDB2RDB Transformation has been used to process the Nomenclátor Geográfico Básico de España (NGBE) dataset. It is available for download at the *Centro de Descargas*[7] of CNIG in one single MS Access file that contains the official gazetteer for Spain. As a result, about 1000000 features have been extracted belonging to 58 different feature types.

The execution of both transformations, MDB2RDB and SHP2RDB, requires executing first a process in order to fill the information in the `DATASETS` and `DATASETS_TABLES_REL` tables.

2.3 From a spatial database to a RDF store

As seen in Section 2, each dataset is stored in the PostgreSQL/PostGIS database as an individual table with some common attributes and other additional ones. These tables have been processed using Morph-RDB in order to obtain a RDF representation of them. Morph-RDB [7], [12] (formerly called ODEMapster [5], [13]) is a RDB2RDF engine developed by the UPM-OEG, which follows the R2RML [8], [14] specification. Figure 6 shows a representation of the model produced by the transformation.

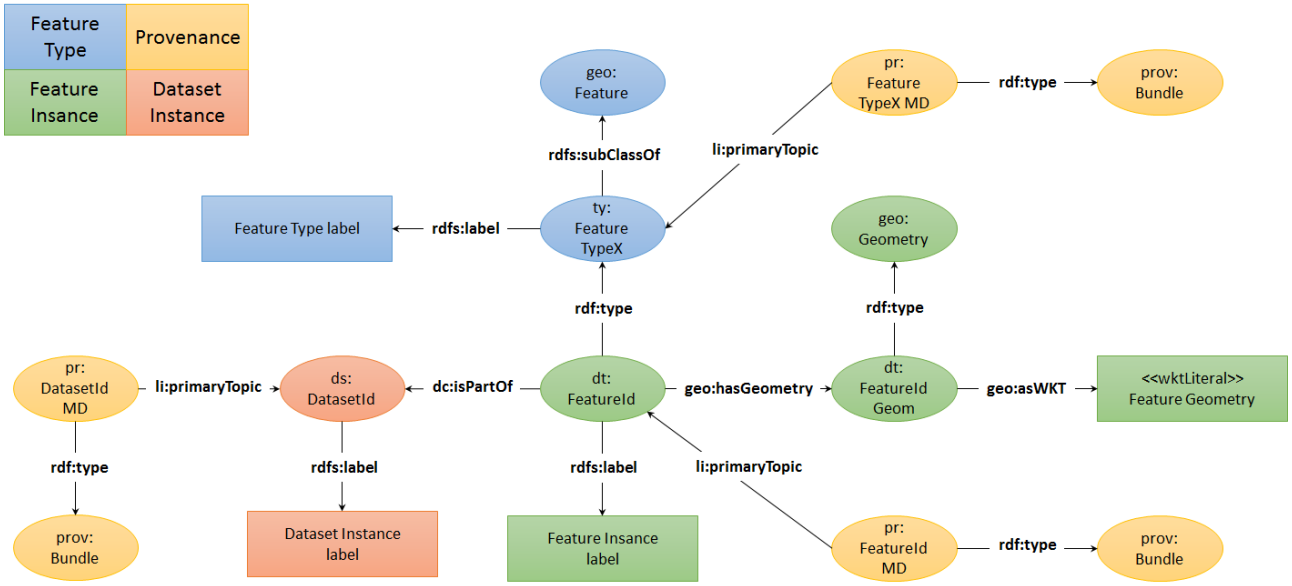

Figure 6 – RDF model

The transformation defines that each feature is an instance of a feature type that in turn is a subclass of the class `geo:Feature`. Each feature has a textual description (`rdfs:label`). Using the `geo:hasGeometry` property each feature points to an entity of type `geo:Geometry` that describes its location. Geometries have a `geo:asWKT` property in which the geometry will be expressed in WKT form. WKT is a plain text serialization format for geometries standardized by OGC and ISO [10], [15]. Provenance

[7] http://centrodedescargas.cnig.es/CentroDescargas/

information is described using properties described in D16.2 *"Call 2: Linked Map VGI Provenance schema"* (`li:primaryTopic`, `prov:Bundle`). Additionally each feature has a relation `dct:isPartOf` to the dataset to which it belongs.

In the listing bellow, an example of extracted feature is attached:

```
dt:BCN-14361923 a geo:Feature ;
  rdfs:label "San Vitero" ;
  dct:isPartOf ds:BCN .
  li:primaryTopic pr:14361923 ;
  geo:hasGeometry [
    a geo:Geometry ;
    geo:asWKT "<http://www.opengis.net/def/crs/OGC/1.3/CRS84>MULTIPOLYGON
(((-0.348315741908539 41.7817277176387, (…) 0.348315741908539 41.78172771
76387)))"^^geo:wktLiteral> ;
  ]
.

pr:14361923 a prov:Bundle ;
  prov:value "<prov:document (…)>(…)</prov:document>"^^rdf:XMLLiteral
.
```

Appendix III contains one of the R2RML files used to transform features from BCN25 dataset from PostgreSQL/PostGIS to RDF. Each file contains three *Triples* sections, each one of them is applied to extract a different Entity:

- Features (TriplesMapFeature)
- Geometries (TriplesMapGeometry)
- Provenance Bundles (TriplesMapPrimaryTopic)

Each *Triples* section can be divided into three parts:

- The first section defines the SQL query that will be used to extract all the features from PostgreSQL/PostGIS
- The second section establishes how the identifier of the Feature will be created, which kind of node will be generated and to which class will belong the Feature.
- The third section consists of one of more *predicateObjectMap*, this kind of section maps one or more columns from the database to a RDF property.

Once we have the dataset converted into RDF, the final step is storing this information in the final triple store. This triple store is built using Strabon [16], a semantic spatiotemporal RDF store developed in the context of European FP7 project SemsorGrid4Env (Semantic Sensor Grids for Rapid Application Development for Environmental Management), that extends the well-known RDF store Sesame[8] to manage thematic, spatial and temporal data that is stored in the backend repository. In order to insert Turtle RDF files in Strabon, we have use the "`eu.earthobservatory.runtime.postgis.StoreOp`" Java utility that is included in the Strabon distribution.

First dumps of the dataset will be available at http://linkedmap.unizar.es/deliverables/ along with the source code employed in their production.

[8] http://www.openrdf.org/

3 Linking system

The linking system is responsible for discovering and establishing the links between the features of different datasets. For this purpose, we have taken advantage of an existing link discovery framework called Silk. Silk is a tool intended to discover relationships between resources from diverse sources. It is open source software that can be run from command line. It additionally offers a web interface that makes easier the configuration and running jobs for establishing links among RDF data.

3.1 Namespaces

The following namespaces and prefixes will be used in this section:

```
geo:     http://www.opengis.net/ont/geosparql#
geof:    http://www.opengis.net/def/function/geosparql/
w3geo:   http://www.w3.org/2003/01/geo/wgs84_pos#
dbpedia: http://es.dbpedia.org/resource/
owl:     http://www.w3.org/2002/07/owl#
dt:      http://linkedmap.unizar.es/id/data/
```

3.2 Silk configuration

Silk projects have three different configuration units:

- Data sources: These elements are the data input of the pairing tasks, they could be whether a SPARQL point (accessible via the Internet) or a RDF dump file ("RDF/XML", "N-TRIPLE", "TURTLE", "TTL", "N3")
- Pairing tasks: In these tasks data sources are paired, each task defines two data sources. It is possible to apply SPARQL restrictions to each data source in order to limit the number of resources.
- Output: It is also possible to select the pairing output, this could be a RDF dump file or a SPARQL Update sentence.

In this context, taking into account that BCN25, NGBE and OSM data will be stored in our SPARQL endpoint, we will need to define two different data sources in Silk:

- BCN25, NGBE and OSM will share the same SPARQL endpoint deployed in a development machine.
- DBpedia SPARQL Live Endpoint[9].

3.3 Pairing tasks

Once data sources are defined, it is necessary to establish the pairing tasks aimed for this project:

- BCN25-NGBE pairing task.
- BCN25-OSM pairing task.
- BCN25-DBpedia pairing task.

Restrictions mentioned in pairing tasks are particularly important in this context for many reasons but the main ones are size matter (the more enclosed the restriction is, the smoother Silk will work) and accuracy (the more precise the restriction is, the less false positives will be found).

BCN25 Dataset is divided into tiles (Figure 7). We will make use of this spatial division in order to minimize the range of resources selected for each task. Having this in mind, each pairing task will be launched once per each tile which will lead us to light task with maximum accuracy. Appendix IV contains an example of pairing file.

[9] http://live.dbpedia.org/sparql

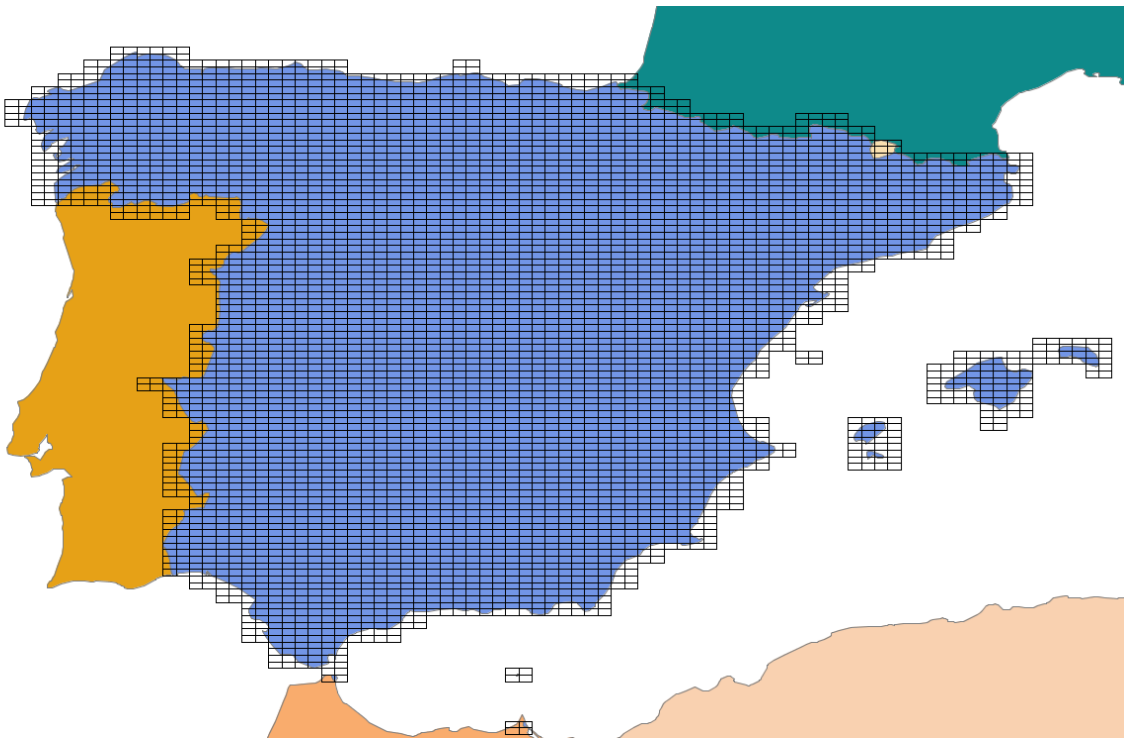

Figure 7 – BCN25 Tile Partitions

For example, in the next listing shows a pairing filter for BCN25-DBpedia:

```
{
  #BCN25 data source restriction:
  ?b geo:hasGeometry ?fGeom.
  ?fGeom geo:asWKT ?y .
  FILTER ( geof:sfIntersects(?y,"$TILEGEOMETRY$") ) .
}

{
  #DBpedia data source restriction:
  ?a w3geo:lat ?lat .
  ?a w3geo:long ?long .
  FILTER(
    ?long >=$MINTILELONG$ &&
    ?long <= $MAXTILELONG$ &&
    ?lat >= $MINTILELAT$ &&
    ?lat <=$MAXTILELAT$).
}
```

As BCN25 will be served in a GeoSPARQL endpoint, we will make use of `sfIntersects` function. The variable `$TILEGEOMETRY$` in `sfIntersects` will be replaced with the BCN25 tile's geometry. We cannot use GeoSPARQL functions in queries to the DBpedia endpoint, so geographic restrictions will be reduced to the tile's bounding box (`$MINTILELONG$` will be replaced with the minimum longitude of the tile, `$MAXTILELONG$` will be the tile's maximum longitude, `$MINTILELAT$` will represent the minimum latitude of the tile and `$MAXTILELAT$` will be replaced with the tile's maximum latitude).

Finally, inside of each pairing task, it has to be defined which labels of each data source will be compared in order to decide if two resources are related.

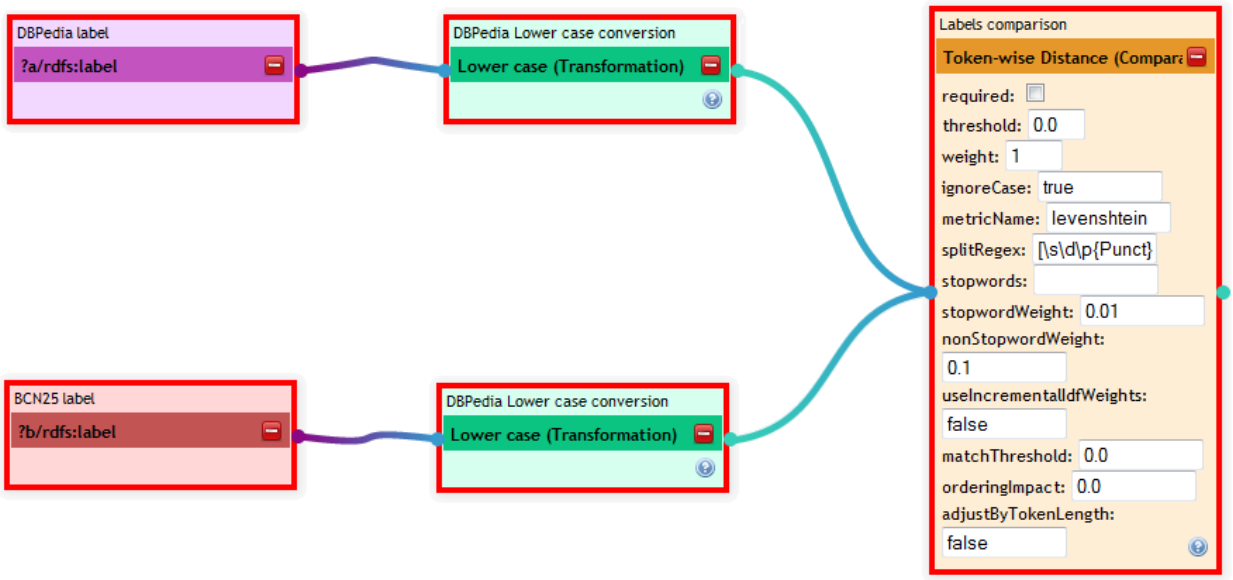

Figure 8 – BCN25-DBPedia comparison task

Figure 8 shows the mapping made between BCN25 and DBpedia resources, in which `rdfs:label` from DBpedia and BCN25 are converted into lower case and then compared using a token-wise distance algorithm. There are many other comparison algorithms but the token-wise distance algorithm is the more suitable in Linked Map context because the Token-wise function splits input in words using (`splitRegex` regular expression) and then compares words using the Levenshtein distance [17].

Finally, as an example, the listing below shows a pairing extract as result of a pairing job obtained with Silk:

```
dbpedia:Campogrande_de_Aliste      owl:sameAs dt:BCN-14361929 .
dbpedia:Bercianos_de_Aliste        owl:sameAs dt:BCN-14361930 .
dbpedia:San_Vicente_de_la_Cabeza   owl:sameAs dt:BCN-14361924 .
dbpedia:Villarino_de_Cebal         owl:sameAs dt:BCN-14361927 .
dbpedia:Palazuelo_de_las_Cuevas    owl:sameAs dt:BCN-14361928 .
dbpedia:Fradellos                  owl:sameAs dt:BCN-14361926 .
dbpedia:Grisuela                   owl:sameAs dt:BCN-14361925 .
dbpedia:San_Vitero                 owl:sameAs dt:BCN-14361923 .
```

4 Conclusions

In this document, we have described the provisioning of an official or governmental geographic dataset (that in the context of the project is called authoritative dataset) and several VGI sources, their transformation into RDF data and their integration using RDF links. For this purpose, we have selected as authoritative geographic dataset the Spanish Numerical Cartographic Database (BCN25/BTN25), which can be downloaded free for non-commercial use and the VGI datasets includes the OSM dataset and the entries with geographic data of Wikipedia.

In order to create the final dataset, two services have been developed and put into practice: the provisioning service and the linking service. The provisioning service was already described in the previous deliverable D16.2 "*Call 2: Linked Map Provisioning service*". This deliverable describes the results obtained after its execution. The linking service, which is also outlined in this document, links entities between datasets based on core gazetteer properties: names, types, location. This service makes use of Silk, an open source tool for discovering relationships between data from different sources.

First dumps of the produced dataset will be available at Linked Map web site[10], along with the source code used in their production[11].

[10] http://linkedmap.unizar.es/datasets/
[11] http://linkedmap.unizar.es/technology/

Appendix I BCN25/BTN25 feature type selection

The table below lists all feature types contained in the BCN25/BTN25 that have been selected to be processed as Linked Data.

Table 1 – BCN25/BTN25 feature types selection

Feature Types	Name	Description
acuedu	Aqueduct	Elevated structure for water conduction
agr_edi	Building group	Group of buildings and facilities with a fixed aim. Usually gathered and organized with a single name.
arbol	Tree	Tree
autop_autov	Motorway/highway	Roads with more than one lane per direction
Calle	Street	Public roads inside of a city
Camino	Pat	Unpaved road
car_bic	Cycle lane	Cycle lane
carr_conv	Conventional Road	Conventional road
Cascade	Waterfall	Waterfall
cau_art	Artificial channel	Human structure to conduct water
Cement	Graveyard	Piece of land devoted to allocate graves
cen_elec, cen_elec_ant_bcn	Power Station	Power Station
con_hid_p, con_hid_s	Hydrographic Structure	Any Structure related to water conduction or usage
con_his_l, con_his_s	Historic Structure	Any Structure considered of historical interest
Cueva	Cave	Underground hollow
edi_ant_bcn, edific	Building	Building
edi_rel, edi_rel_ant_bcn	Religious Building	Building devoted to religious purposes
eje_con	Hydrographic Axis	Imaginary line describing a river
ent_pob	Settlement	Human settlement
est_fc	Rail station	Rail station
exp_min_p	Mining camp	Point or Surface devoted to minery
inf_ffcc_ant_bcn	Railway Facilities	Any railway structure
ins_dep	Sport Facilities	Any sport structure
ins_ind	Industrial Facilities	Any industrial structure
ins_rec_p, ins_rec_s	Leisure Facilities	Any leisure structre
Isla	Island	Island
itiner, senda	Itinerary	Route
kil_carr	Road milestone	Kilometric signal in a road

Feature Types	Name	Description
Laguna, lin_emb	Lake	Lake
lin_rot	Mountain Chain	Imaginary line describing a mountain chain
mol_ant_bcn, Molino	Mill	Building devoted to mill
monume	Monument	Building considered of historial interest
Presa	Reservoir	Human structure in a water flow used to reservoir water
Puerto	Harbour	Harbour
pun_aco	Reference point	Reference point with a known altitude used to interpret maps
Rambla, ria, rio	Water flow	Natural water flow
Salina	Salt lake	Lake with salted water
top_pos_geo, top_sin_geo	Place name	Place name

Appendix II Example of PROV-XML template

The template below is an example of the PROV-XML templates used for gathering provenance information during the SHP2RDB transformation process. The template has the following variables:

- `ID_GEOKETTLE`: The identifier of the GeoKettle version that run the transformation.
- `LOCAL_FILE_URI`: The identifier of the Esri shapefile used in the generation.
- `LOCAL_FILE_TITLE`: A descriptive title of the entity used.
- `LOCAL_FILE_LOCATION`: The location of the Esri shapefile used.
- `LOCAL_ID`: The identifier of the row in the Esri shapefile from which the data is extracted.
- `UNIQUE_ID`: The identifier of the generated feature.
- `RUN_ID`: The identifier of the run.
- `RUN_START_DATE`: The start date of the run.
- `GEOKETTLE_TRANSFORMATION_ID`: The identifier of the GeoKettle Transformation run.
- `GEOKETTLE_TRANSFORMATION_TITLE`: A descriptive title of the GeoKettle Transformation.
- `GEOKETTLE_TRANSFORMATION_LOCATION`: The location of the GeoKettle Transformation.

```
<?xml version="1.0" encoding="UTF-8"?>
<prov:document
  xmlns:li="http://linkedmap.unizar.es/def/prov-li#"
  xmlns:ds="http://linkedmap.unizar.es/id/dataset/"
  xmlns:dt="http://linkedmap.unizar.es/id/data/"
  xmlns:geo="http://www.opengis.net/ont/geosparql#"
  xmlns:dct="http://purl.org/dc/dcmitype/"
  xmlns:prov="http://www.w3.org/ns/prov#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">

  <prov:softwareAgent prov:id="${ID_GEOKETTLE}" />

  <prov:entity prov:id="ds:${LOCAL_FILE_URI}">
    <prov:label>${LOCAL_FILE_TITLE}</prov:label>
    <prov:location
xsi:type="xsd:anyURI">${LOCAL_FILE_LOCATION}</prov:location>
    <prov:type xsi:type="xsd:QName">dct:Dataset</prov:type>
  </prov:entity>

  <prov:entity prov:id="ds:${LOCAL_FILE_URI}#${LOCAL_ID}">
    <prov:label>Feature at ${LOCAL_FILE_TITLE}</prov:label>
    <prov:type xsi:type="xsd:QName">geo:Feature</prov:type>
  </prov:entity>

  <prov:hadMember>
    <prov:collection prov:ref="ds:${LOCAL_FILE_URI}"/>
    <prov:entity prov:ref="ds:${LOCAL_FILE_URI}#${LOCAL_ID}"
  </prov:hadMember>
```

```
  <prov:entity prov:id="dt:${UNIQUE_ID}">
    <prov:type xsi:type="xsd:QName">geo:Feature</prov:type>
  </prov:entity>

  <prov:activity prov:id="${RUN_ID}">
    <prov:label>Run GeoKettle Transformation</prov:label>
    <prov:startTime>${RUN_START_DATE}</prov:startTime>
  </prov:activity>

  <prov:plan prov:id="${GEOKETTLE_TRANSFORMATION_ID}">
    <prov:label>${GEOKETTLE_TRANSFORMATION_TITLE}</prov:label>
    <prov:location
xsi:type="xsd:anyURI">${GEOKETTLE_TRANSFORMATION_LOCATION}</prov:location
>
  </prov:plan>

  <prov:wasAssociatedWith>
    <prov:activity prov:ref="${RUN_ID}"/>
    <prov:agent prov:ref="${ID_GEOKETTLE}"/>
    <prov:plan prov:ref="${GEOKETTLE_TRANSFORMATION_ID}"/>
  </prov:wasAssociatedWith>

  <prov:wasDerivedFrom>
    <prov:generatedEntity prov:ref="dt:${UNIQUE_ID}"/>
    <prov:usedEntity prov:ref="ds:${LOCAL_FILE_URI}#${LOCAL_ID}"/>
    <prov:activity prov:ref="${RUN_ID}"/>
  </prov:wasDerivedFrom>

</prov:document>
```

Appendix III Example of R2RML File

R2RML definition used for the extraction of the features stored in the table ENT_POB that correspond with instances of the feature type "*Entidad Poblacion*".

```
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ex: <http://example.com/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rev: <http://purl.org/stuff/rev#> .
@prefix bsbm:
  <http://www4.wiwiss.fu-berlin.de/bizer/bsbm/v01/vocabulary/> .
@prefix bsbm-inst:
  <http://www4.wiwiss.fu-berlin.de/bizer/bsbm/v01/instances/> .
@base <http://mappingpedia.org/rdb2rdf/r2rml/tc/> .

<TriplesMapFeature> a rr:TriplesMap;

  rr:logicalTable [
    rr:sqlQuery """
      SELECT ent_pob.id AS id,
             ent_pob.label AS label,
             ent_pob.project AS project
      FROM ent_pob"""
  ];

  rr:subjectMap [
    a rr:Subject;
      rr:template "http://linkedmap.unizar.es/data/{project}-{id}";
      rr:termType rr:IRI;
      rr:class geo:Feature;
  ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant rdfs:label;
    ];
    rr:objectMap [
      rr:termType rr:Literal;
      rr:column "label";
    ];
  ];

  rr:predicateObjectMap [
    rr:predicate dct:isPartOf;
    rr:objectMap [
      rr:template "http://linkedmap.unizar.es/data/{project}"
    ];
  ];
```

```
  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant geo:hasGeometry ;
    ];
    rr:objectMap    [
      rr:parentTriplesMap <TriplesMapGeometry>;
      rr:joinCondition [
        rr:child "id" ;
        rr:parent "id" ;
      ] ;
    ];
  ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant li:primaryTopic;
    ];
    rr:objectMap [
      rr:parentTriplesMap <TriplesPrimaryTopic>;
      rr:joinCondition [
        rr:child "id" ;
        rr:parent "id" ;
      ]
    ];
  ];
.

<TriplesPrimaryTopic>
  a rr:TriplesMap;

  rr:logicalTable [
    rr:sqlQuery  """
      SELECT ent_pob.id AS id,
             ent_pob.prov AS prov,
             ent_pob.project AS project
      FROM ent_pob"""
  ];

  rr:subjectMap [
    a rr:Subject;
    rr:template "http://linkedmap.unizar.es/id/provs/{id}";
    rr:termType rr:IRI;
    rr:class prov:Bundle;
  ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant prov:value
    ];
    rr:objectMap [
      rr:termType rr:Literal;
      rr:column "prov";
    ];
  ];
.
```

```
<TriplesMapGeometry>
  a rr:TriplesMap;

  rr:logicalTable [
    rr:sqlQuery  """
      SELECT ent_pob.id AS id ,
             ST_AsText(ent_pob.geom) as geometry FROM ent_pob"""
  ];

  rr:subjectMap [
    a rr:Subject;
    rr:template "geom{id}";
    rr:termType rr:BlankNode;
    rr:class geo:Geometry;
  ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant geo:asWKT
    ];
    rr:objectMap [
      rr:termType rr:Literal;
      rr:template
         "<http://www.opengis.net/def/crs/OGC/1.3/CRS84>{geometry}";
      rr:datatype geo:wktLiteral
    ];
  ];
.
```

Appendix IV Silk project Example

In this configuration file DBpedia and BCN25 data source are selected. Labels that will be paired are `rdfs:label` from both sources will be converted into lower case and then compared with a `TokenWise Comparator` provided by Silk.

```
<?xml version="1.0" encoding="UTF-8"?>
<Silk>
  <Prefixes>
    <Prefix namespace="http://www.opengis.net/ont/geosparql#"
      id="geostrabon" />
    <Prefix namespace="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      id="rdf" />
    <Prefix namespace="http://www.w3.org/2003/01/geo/wgs84_pos#"
      id="geo" />
    <Prefix namespace="http://www.w3.org/2002/07/owl#" id="owl" />
    <Prefix namespace="http://www.w3.org/2000/01/rdf-schema#"
      id="rdfs" />
    <Prefix namespace="http://www.opengis.net/def/function/geosparql/"
      id="geof" />
    <Prefix namespace="http://www.opengis.net/ont/geosparql#"
      id="geosparql" />
  </Prefixes>
  <DataSources>
    <DataSource type="sparqlEndpoint" id="esDBpedia">
      <Param name="pageSize" value="1000" />
      <Param name="pauseTime" value="0" />
      <Param name="retryCount" value="3" />
      <Param name="endpointURI" value="http://es.dbpedia.org/sparql" />
      <Param name="retryPause" value="1000" />
      <Param name="graph" value="" />
      <Param name="login" value="" />
      <Param name="entityList" value="" />
      <Param name="password" value="" />
      <Param name="additionalParameters" value="" />
    </DataSource>
    <DataSource type="sparqlEndpoint" id="strabon">
      <Param name="pageSize" value="1000" />
      <Param name="pauseTime" value="0" />
      <Param name="retryCount" value="3" />
      <Param name="endpointURI"
        value="http://descartes.cps.unizar.es:8080/strabonendpoint/Query"
/>
      <Param name="retryPause" value="1000" />
      <Param name="graph" value="" />
      <Param name="login" value="" />
      <Param name="entityList" value="" />
      <Param name="password" value="" />
      <Param name="additionalParameters" value="" />
    </DataSource>
  </DataSources>
```

```
  <Interlinks>
    <Interlink id="linkFeatures">
      <LinkType>owl:sameAs</LinkType>
      <SourceDataset dataSource="esDBpedia" var="a">
        <RestrictTo> ?a
          &lt;http://www.w3.org/2003/01/geo/wgs84_pos#lat&gt;
          ?lat .
        </RestrictTo>
      </SourceDataset>
      <TargetDataset dataSource="strabon" var="b">
        <RestrictTo>
          ?b
          &lt;http://www.opengis.net/ont/geosparql#hasGeometry&gt;
          ?fGeom.
          ?fGeom &lt;http://www.opengis.net/ont/geosparql#asWKT&gt; ?y .
          FILTER (
          &lt;http://www.opengis.net/def/function/geosparql/sfIntersects
          &gt;(?y,&quot;MULTIPOLYGON(((-6.3541174445282
          41.8334119003174,-6.18745112984291
          41.8334139949847,-6.18745204026853
          41.7500792156347,-6.35411835196593
          41.7500799015362,-6.3541174445282 41.8334119003174)))&quot;)) .
        </RestrictTo>
      </TargetDataset>
      <LinkageRule>
        <Compare weight="1" threshold="0.0" required="false"
          metric="tokenwiseDistance" id="unnamed_6">
          <TransformInput function="lowerCase" id="unnamed_3">
            <Input path="?a/rdfs:label" id="unnamed_1" />
          </TransformInput>
          <TransformInput function="lowerCase" id="unnamed_4">
            <Input path="?b/rdfs:label" id="unnamed_2" />
          </TransformInput>
          <Param name="splitRegex" value="[\s\d\p{Punct}]+" />
          <Param name="metricName" value="levenshtein" />
          <Param name="orderingImpact" value="0.0" />
          <Param name="useIncrementalIdfWeights" value="false" />
          <Param name="adjustByTokenLength" value="false" />
          <Param name="stopwordWeight" value="0.01" />
          <Param name="nonStopwordWeight" value="0.1" />
          <Param name="ignoreCase" value="true" />
          <Param name="matchThreshold" value="0.0" />
        </Compare>
      </LinkageRule>
      <Filter />
      <Outputs />
    </Interlink>
  </Interlinks>
</Silk>
```

Appendix V Geokettle configuration file

GeoKettle configuration file


```
<?xml version="1.0" encoding="UTF-8"?>
<transformation>
  <info>
    <name>ent_pob</name>
    <description />
    <extended_description />
    <trans_version />
    <trans_status>0</trans_status>
    <directory>/</directory>
    <parameters>
      <parameter>
        <name>shapeFileName</name>
        <default_value>null</default_value>
        <description>null</description>
      </parameter>
    </parameters>
    <log>
      <read />
      <write />
      <input />
      <output />
      <update />
      <rejected />
      <connection />
      <table />
      <step_performance_table />
      <use_batchid>Y</use_batchid>
      <use_logfield>N</use_logfield>
      <size_limit_lines />
    </log>
    <maxdate>
      <connection />
      <table />
      <field />
      <offset>0.0</offset>
      <maxdiff>0.0</maxdiff>
    </maxdate>
    <size_rowset>10000</size_rowset>
    <sleep_time_empty>50</sleep_time_empty>
    <sleep_time_full>50</sleep_time_full>
    <unique_connections>N</unique_connections>
    <feedback_shown>Y</feedback_shown>
    <feedback_size>50000</feedback_size>
    <using_thread_priorities>Y</using_thread_priorities>
    <shared_objects_file />
    <capture_step_performance>N</capture_step_performance>
    <step_performance_capturing_delay>1000
    </step_performance_capturing_delay>
```

```
    <dependencies />
    <partitionschemas />
    <slaveservers />
    <clusterschemas />
    <modified_user>-</modified_user>
    <modified_date>2014/03/24 09:15:07.667</modified_date>
  </info>
  <notepads>
    <notepad>
      <note>E:\GeoKettle&gt;Pan.bat
        -file:"e:\planetData\ent_pob.ktr" level:Basic
        -
param:"shapeFileName=E:\planetData\shps\BCN25_2373\BCN0501S_ENT_POB.shp"
      </note>
      <xloc>124</xloc>
      <yloc>44</yloc>
      <width>823</width>
      <heigth>22</heigth>
    </notepad>
  </notepads>
  <connection>
    <name>descartesBCNBTN</name>
    <server>descartes</server>
    <type>POSTGRESQL</type>
    <access>Native</access>
    <database>bcnbtn</database>
    <port>5432</port>
    <username>postgres</username>
    <password>Encrypted 2be98afc86aa7f2e4bb16bd64d980aac9</password>
    <servername />
    <data_tablespace />
    <index_tablespace />
    <attributes>
      <attribute>
        <code>FORCE_IDENTIFIERS_TO_LOWERCASE</code>
        <attribute>N</attribute>
      </attribute>
      <attribute>
        <code>FORCE_IDENTIFIERS_TO_UPPERCASE</code>
        <attribute>N</attribute>
      </attribute>
      <attribute>
        <code>IS_CLUSTERED</code>
        <attribute>N</attribute>
      </attribute>
      <attribute>
        <code>PORT_NUMBER</code>
        <attribute>5432</attribute>
      </attribute>
      <attribute>
        <code>QUOTE_ALL_FIELDS</code>
        <attribute>N</attribute>
      </attribute>
      <attribute>
        <code>SUPPORTS_BOOLEAN_DATA_TYPE</code>
        <attribute>N</antribute>
      </attribute>
```

```
      <attribute>
        <code>USE_POOLING</code>
        <attribute>N</attribute>
      </attribute>
    </attributes>
  </connection>
  <order>
    <hop>
      <from>SRS Transformation</from>
      <to>Salida Tabla</to>
      <enabled>Y</enabled>
    </hop>
    <hop>
      <from>Shapefile File Input</from>
      <to>Valor Java Script Modificado</to>
      <enabled>Y</enabled>
    </hop>
    <hop>
      <from>Valor Java Script Modificado</from>
      <to>SRS Transformation</to>
      <enabled>Y</enabled>
    </hop>
  </order>
  <step>
    <name>SRSTransformation</name>
    <type>SRSTransformation</type>
    <description />
    <distribute>Y</distribute>
    <copies>1</copies>
    <partitioning>
      <method>none</method>
      <schema_name />
    </partitioning>
    <field_name>the_geom</field_name>
    <source_gui_status>3</source_gui_status>
    <target_gui_status>1</target_gui_status>
    <source_srs>
      <authority>Custom Authority</authority>
      <srid>-1</srid>
      <description>Custom SRS from WKT</description>
      <is_custom>Y</is_custom>
      <wkt>PROJCS["ETRS_1989_UTM_Zone_29N",
        GEOGCS["GCS_ETRS_1989",
        DATUM["D_ETRS_1989",
        SPHEROID["GRS_1980", 6378137.0, 298.257222101]],
        PRIMEM["Greenwich", 0.0],
        UNIT["degree", 0.017453292519943295],
        AXIS["Longitude", EAST],
        AXIS["Latitude", NORTH]],
        PROJECTION["Transverse_Mercator"],
        PARAMETER["central_meridian", -9.0],
        PARAMETER["latitude_of_origin", 0.0],
        PARAMETER["scale_factor", 0.9996],
        PARAMETER["false_easting", 500000.0],
        PARAMETER["false_northing", 0.0],
        UNIT["m", 1.0],
        AXIS["x", EAST],
```

```
        AXIS["y", NORTH]]
      </wkt>
    </source_srs>
    <target_srs>
      <authority>EPSG</authority>
      <srid>4258</srid>
      <description>ETRS89</description>
      <is_custom>N</is_custom>
    </target_srs>
    <cluster_schema />
    <remotesteps>
      <input />
      <output />
    </remotesteps>
    <GUI>
      <xloc>476</xloc>
      <yloc>178</yloc>
      <draw>Y</draw>
    </GUI>
  </step>
  <step>
    <name>Database Updater</name>
    <type>TableOutput</type>
    <description />
    <distribute>Y</distribute>
    <copies>1</copies>
    <partitioning>
      <method>none</method>
      <schema_name />
    </partitioning>
    <connection>descartesBCNBTN</connection>
    <schema />
    <table>ent_pob</table>
    <commit>1000</commit>
    <truncate>N</truncate>
    <ignore_errors>N</ignore_errors>
    <use_batch>Y</use_batch>
    <specify_fields>Y</specify_fields>
    <partitioning_enabled>N</partitioning_enabled>
    <partitioning_field />
    <partitioning_daily>N</partitioning_daily>
    <partitioning_monthly>Y</partitioning_monthly>
    <tablename_in_field>N</tablename_in_field>
    <tablename_field />
    <tablename_in_table>Y</tablename_in_table>
    <return_keys>N</return_keys>
    <return_field />
    <fields>
      <field>
        <column_name>id</column_name>
        <stream_name>ID</stream_name>
      </field>
      <field>
        <column_name>etiqueta</column_name>
        <stream_name>ETIQUETA</stream_name>
      </field>
      <field>
```

```
        <column_name>proyecto</column_name>
        <stream_name>PROYECTO</stream_name>
      </field>
      <field>
        <column_name>geom</column_name>
        <stream_name>the_geom</stream_name>
      </field>
      <field>
        <column_name>prov</column_name>
        <stream_name>prov</stream_name>
      </field>
      <field>
        <column_name>tipo</column_name>
        <stream_name>TIPO_0501</stream_name>
      </field>
    </fields>
    <cluster_schema />
    <remotesteps>
      <input />
      <output />
    </remotesteps>
    <GUI>
      <xloc>675</xloc>
      <yloc>180</yloc>
      <draw>Y</draw>
    </GUI>
  </step>
  <step>
    <name>ShapeFile Reader</name>
    <type>GISFileInput</type>
    <description />
    <distribute>Y</distribute>
    <copies>1</copies>
    <partitioning>
      <method>none</method>
      <schema_name />
    </partitioning>
    <filename>${shapeFileName}</filename>
    <isfilenameinfield>N</isfilenameinfield>
    <filenamefield />
    <limit>0</limit>
    <add_rownr>N</add_rownr>
    <field_rownr />
    <gis_file_charset>UTF-8</gis_file_charset>
    <passing_through_fields>N</passing_through_fields>
    <accept_stepname />
    <cluster_schema />
    <remotesteps>
      <input />
      <output />
    </remotesteps>
    <GUI>
      <xloc>189</xloc>
      <yloc>176</yloc>
      <draw>Y</draw>
    </GUI>
  </step>
```

```
  <step>
    <name>Provenance appender</name>
    <type>ScriptValueMod</type>
    <description />
    <distribute>Y</distribute>
    <copies>1</copies>
    <partitioning>
      <method>none</method>
      <schema_name />
    </partitioning>
    <compatible>N</compatible>
    <jsScripts>
      <jsScript>
        <jsScript_type>0</jsScript_type>
        <jsScript_name>Script 1</jsScript_name>
        <jsScript_script>//Script here
          var prov="provenance"
        </jsScript_script>
      </jsScript>
    </jsScripts>
    <fields>
      <field>
        <name>prov</name>
        <rename>prov</rename>
        <type>String</type>
        <length>-1</length>
        <precision>-1</precision>
        <replace>N</replace>
      </field>
    </fields>
    <cluster_schema />
    <remotesteps>
      <input />
      <output />
    </remotesteps>
    <GUI>
      <xloc>353</xloc>
      <yloc>284</yloc>
      <draw>Y</draw>
    </GUI>
  </step>
  <step_error_handling />
  <slave-step-copy-partition-distribution />
  <slave_transformation>N</slave_transformation>
</transformation>
```

References

[1] F. J. López-Pellicer and J. Barrera, “D16.2 Call 2: Linked Map Provisioning service,” PlanetData, 2014.

[2] J. Barrera and F. J. López-Pellicer, “D16.4 *Call 2: Linked Map Data access/update service*,” PlanetData.

[3] F. J. López-Pellicer and J. Barrera, “D15.1 *Call 2: Linked Map requirements definition and conceptual architecture*,” PlanetData, 2014.

[4] J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov, “Silk-A Link Discovery Framework for the Web of Data.,” presented at the Workshop on Linked Data on the Web, 2009.

[5] Esri, “ESRI Shapefile Technical Description,” Esri Press, Jul. 1998.

[6] F. J. López-Pellicer and J. Barrera, “D16.1 Call 2: Linked Map VGI provenance schema,” PlanetData, 2014.

[7] T. Badard, “GeoKettle: a powerful open source spatial ETL tool,” presented at the FOSS4G 2009 Conference, Sydney, Australia, 2009.

[8] C. Strobl, “PostGIS,” *Encyclopedia of GIS*, pp. 891–898, 2008.

[9] H. Hua, C. Tilmes, and S. Zednik, Eds., “PROV-XML: The PROV XML Schema,” 30 April 2013, W3C Working Group Note.

[10] A. F. Rodríguez and L. García Asensio, “A fully integrated information system to manage cartographic and geographic data at 1:25.000 scale,” presented at the 22nd International Cartographic Conference, 2005.

[11] F. J. García, A. de las Cuevas, A. Marín, V. Martin, F. Sánchez, and F. J. González-Matesanz, “New production environment for the National Topographic Database 1:25.000 (IGN-E). Intelligence for geographic databases.,” presented at the 26th International Cartographic Conference, 2013.

[12] F. Priyatna, O. Corcho, and J. F. Sequeda, “Formalisation and Experiences of R2RML-based SPARQL to SQL Query Translation using Morph ,” presented at the WWW 2014, 2014.

[13] J. B. Rodriguez and A. Gómez-Pérez, “Upgrading relational legacy data to the semantic web,” *World Wide Web Conference*. ACM, May-2006.

[14] S. Das, S. Sundara, and R. Cyganiak, “R2RML: RDB to RDF Mapping Language,” W3C Recommendation 27 September 2012, 2012.

[15] ISO/IEC, “Information technology -- Database languages -- SQL multimedia and application packages -- Part 3: Spatial,” ISO/IEC 13249-3:2011, 2011.

[16] K. Kyzirakos, M. Karpathiotakis, and M. Koubarakis, “Strabon: A Semantic Geospatial DBMS,” in *Lecture Notes in Computer Science*, vol. 7649, no. 19, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 295–311.

[17] V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions and reversals,” *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, 1966.