

GeoTriples: a Tool for Publishing Geospatial Data as RDF Graphs Using R2RML Mappings

Kostis Kyzirakos¹, Ioannis Vlachopoulos², Dimitrianos Savva²,
Stefan Manegold¹, and Manolis Koubarakis²

¹ Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
{`firstname.lastname`}@`cwi.nl`

² National and Kapodistrian University of Athens, Greece
{`johnv1,dimis,koubarak`}@`di.uoa.gr`

Abstract. In this paper we present the tool GeoTriples that allows the transformation of Earth Observation data and geospatial data into RDF graphs, by using and extending the R2RML mapping language to be able to deal with the specificities of geospatial data. GeoTriples is a semi-automated tool that transforms geospatial information into RDF following the state of the art vocabularies like GeoSPARQL and stSPARQL, but at the same time it is not tightly coupled to a specific vocabulary.

Keywords: Linked Geospatial Data, data publishing, GeoSPARQL, stSPARQL

1 Introduction

In the last few years there has been significant effort on publishing EO and geospatial data sources as linked open data. However, the problem of publishing geospatial data sources into RDF graphs using a generic and extensible framework has received little attention as it has only recently emerged. Instead, scripting methods, that were adapted to the subject, were employed mostly for this task, such as custom python scripts developed in project TELEIOS³. However, some work towards developing automated methods for translating geospatial data into RDF has been presented in the latest LGD Workshop⁴. In this paper we present the tool GeoTriples that allows the transformation of geospatial data stored in spatially-enabled relational databases and raw files. It is implemented as an extension to the D2RQ platform⁵ [1] and goes beyond the state of the art by extending the R2RML mapping language⁶ to deal with the specificities of geospatial data. GeoTriples uses GeoSPARQL⁷ as the target vocabulary but the user is free to use any vocabulary she finds appropriate.

³<http://www.earthobservatory.eu>

⁴<http://www.w3.org/2014/03/lgd>

⁵<http://d2rq.org>

⁶<http://www.w3.org/TR/r2rml/>

⁷<http://www.opengeospatial.org/standards/geosparql/>

| NAME | TYPE | WIDTH |
|-----------|--------|-------|
| Zeitlbach | stream | 1 |
| Mangfall | river | 25 |
| Triftbach | canal | 10 |

```

osm_w:1 rdf:type geo:Feature ;
        osm_ont:hasName "Mangfall"^^xsd:string ;
        geo:hasGeometry osm_g:1 .
osm_g:1 rdf:type geo:Geometry ;
        geo:dimension "2"^^xsd:integer .

```

(a) Example data from an ESRI shapefile (b) Expected RDF triples about Mangfall

```

_:osm
  rr:logicalTable [ rr:tableName "'osm'"; ];
  rr:subjectMap [
    rr:class geo:Feature;
    rr:template "http://data.example.com/osm-waterways/
                Feature/id/{'gid'}"; ];
  rr:predicateObjectMap [
    rr:predicate osm:hasName;
    rr:objectMap [ rr:datatype xsd:string;
                  rr:column "'NAME'"; ]; ];
  rr:predicateObjectMap [
    rr:predicate geo:hasGeometry ;
    rr:objectMap [
      rr:parentTriplesMap _:osmGeometry;
      rr:joinCondition [
        rr:child "gid";
        rr:parent "gid"; ]; ]; ].

_:osmGeometry
  rr:logicalTable [ rr:tableName "'osm'"; ];
  rr:subjectMap [
    rr:class geo:Geometry;
    rr:template "http://data.example.com/osm-waterways/
                Geometry/id/{'gid'}"; ];
  rr:predicateObjectMap [
    rr:predicate geo:dimension;
    rr:objectMap [
      rrx:transformation [
        rrx:function geof:dimension;
        rrx:argumentMap (
          [rr:column "'geom'"]
        ); ]; ]; ].

```

(c) Mapping of thematic information (d) Mapping of geometric information

Fig. 1: Examples of extended R2RML mappings for OSM

2 The Tool GeoTriples

GeoTriples⁸ is an open source tool, that takes as input geospatial data that are stored in a spatially enabled database, data that reside in raw files (e.g. ESRI shapefiles) or the results that derive from processing of the aforementioned data (e.g. a SciQL query over raster or array data). At a lower level, GeoTriples uses a connector for each type of input data that transparently accesses and processes the input data. It also consists of two main components: the mapping generator and the R2RML processor. The mapping generator creates automatically an R2RML mapping document from the input data source. The mapping is also enriched with subject and predicate object maps so that the RDF graph that will be produced follows the GeoSPARQL vocabulary. Geospatial information is modeled using a variety of data models (e.g., relational, hierarchical) and is made available using a variety of formats (e.g., ESRI shapefiles, KML documents). In order to deal with these specificities of geospatial information, we extended the R2RML language to allow the representation of a transformation function over the input data via an object map. In [2] we provide more information about our approach. Figure 1 presents an example of such a transformation. The R2RML processor is responsible for producing the desired RDF output by taking into account the mapping document generated, which is also optionally edited by the user. When the R2RML processor of GeoTriples detects an object map with a transformation function, it applies on the fly this function on the serialization of the geometry described in the subject map. However, if the input data source is a spatially enabled database, it generates the appropriate SQL queries that push these transformations to the underlying DBMS.

⁸<https://sourceforge.net/projects/geotriples/>

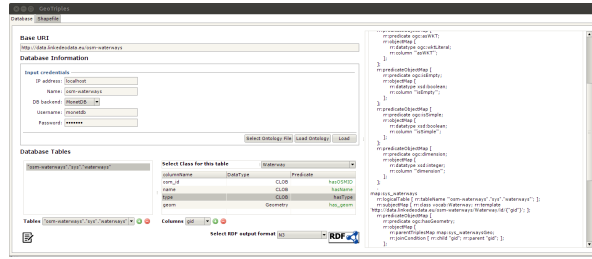


Fig. 2: The graphical user interface of GeoTriples

3 Using GeoTriples in a real-world scenario

In this section we present how we will demonstrate the tool GeoTriples in the context of a precision farming application that is developed by the FP7 EU project LEO⁹. The application combines traditional geospatial data with linked geospatial data for enhancing the quality of precision farming activities. Precision farming aims to solve numerous problems for farmers such as the minimization of the environmental pollution by fertilizers. For dealing with this issue, the farmers have to comply with many legal and technical guidelines that require the combination of information that resides in diverse information sources. In this section we present how linked geospatial data can form the knowledge base for providing solutions for this problem. We will publish the following datasets as RDF graphs using GeoTriples in order to use them in the precision farming application.

OpenStreetMap (OSM) is a collaborative project for publishing free maps of the world. OSM maintains a community-driven global editable map that gathers map data in a crowdsourcing fashion.

Talking Fields aims to increase the efficiency of agricultural production via precision farming by means of geo-information services integrating space and ground-based assets. It produces products for improved soil probing using satellite-based zone maps, and provide services for monitoring crop development through provision of biomass maps and yield estimates.

Natura 2000 is a European ecological network where national authorities submit a standard data form that describes each site and its ecology in order to be characterized as a Natura site.

Corine Land Cover (CLC) is an activity of the European Environment Agency that collects data regarding the land cover of European countries.

In this demo we will use GeoTriples in order to produce the R2RML mappings that dictate the process of generating the desired RDF output from the above data. Then, using the R2RML processor of GeoTriples, we translate the input data into RDF graphs and store the latter into the geospatial RDF store Strabon¹⁰ [3].

⁹<http://www.linkedeodata.eu>

¹⁰<http://www.strabon.di.uoa.gr/>

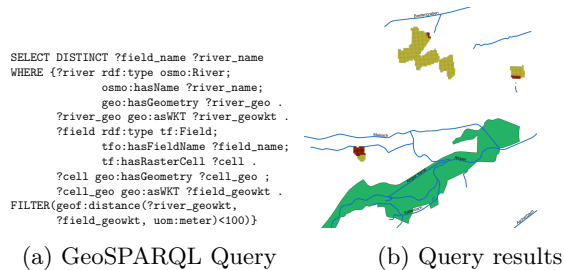


Fig. 3: Discover the parts of the agricultural fields that are close to rivers.

The user can use the graphical interface of GeoTriples that is displayed in Figure 2 for publishing these datasets as RDF graphs. At first the user defines the necessary credentials of the DBMS that stores the above datasets. Then, she selects the tables and the columns that contain the information that she want to publish as RDF graphs. Optionally, an existing ontology may be loaded, in order to map the columns of the selected table to properties from the loaded ontology and map the instances generated from the rows to a specific class. Afterwards, GeoTriples generates automatically the R2RML mappings and presents them to the user. Finally, the user may either customize the generated mappings or proceed with the generation of the RDF graph.

A series of GeoSPARQL queries will be issued afterwards in Strabon for providing the precision farming application with information like the location of agricultural fields that are close to a river. This information allows the precision farming application to take into account legal restrictions regarding distance requirements when preparing the prescription maps that the farmers will utilize afterwards. In Figure 3a we present a GeoSPARQL query that discovers this information, and in Figure 3b we depict the query results.

4 Conclusions

In this paper we presented the tool GeoTriples that uses an extended form of R2RML mapping language to transform geospatial data into RDF graphs, and the GeoSPARQL ontology to properly express it. We demonstrate how GeoTriples is being used for publishing geospatial information that resides in different data sources for the realization of a precision farming application.

References

1. C. Bizer and A. Seaborne. D2RQ-treating non-RDF databases as virtual RDF graphs. In *Proceedings of the 3rd International Semantic Web Conference*, 2004.
2. K. Kostis, V. Ioannis, S. Dimitrianos, M. Stefan, and K. Manolis. Data models and languages for mapping EO data to RDF. Del. 2.1, FP7 project LEO, 2014.
3. K. Kyzirakos, M. Karpathiotakis, and M. Koubarakis. Strabon: A Semantic Geospatial DBMS. In *International Semantic Web Conference*, 2012.