1 LinkedGeoData

1.1 Introduction

The goal of the LinkedGeoData[1] [1] [2] (LGD) project is to ease data integration tasks which require spatial background knowledge. Technically, this is realized by lifting OpenStreetMap[2] (OSM) data into the Semantic Web infrastructure by converting it to RDF[3], interlinking it with other Semantic Web knowledge bases, and finally providing access to all obtained data via SPARQL[4] and Linked Data[5].

The declared goal of OSM is to *create a free map of the world.* So far, its community succeeded in creating a huge geospatial knowledge base built from the following primitives:

- *Nodes:* These are point geometries with latitude/longitude information. Currently, there are 1.64 billion nodes in the OSM database.
- *Ways:* Sequences of nodes. At present, there exist about 156 million ways.
- *Relations:* Aaggregations of nodes, ways and relations, which are used for instance to model multipolygons. There are about 1.64 millions relations.

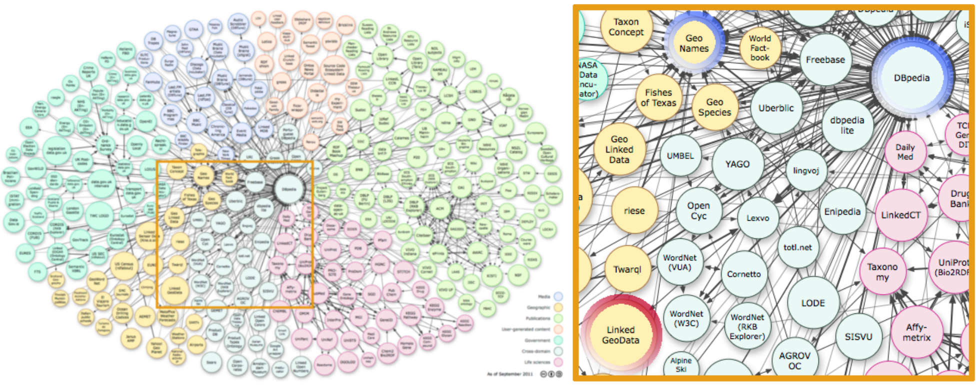
Each of this primitives is described using a set of key-value pairs, referred to as *tags*.

In the course of the LinkedGeoData project, this data was first converted to RDF and afterwards interlinked with DBpedia[6] and GeoNames[7]. LinkedGeoData within the Linked Open Data (LOD) cloud is depicted in Figure 1.1. It is one of the largest existing RDF datasets.

1.2 Architecture

The present architecture of LGD is depicted in Figure 1.2 and will subsequently be briefly explained. OSM provides both full and incremental dumps (minutely,

[1] `http://linkedgeodata.org`
[2] `http://openstreetmap.org`
[3] `http://www.w3.org/TR/2004/REC-rdf-primer-20040210/`
[4] `http://www.w3.org/TR/sparql11-query/`
[5] `http://www.w3.org/DesignIssues/LinkedData.html`
[6] `http://dbpedia.org`
[7] `http://geonames.org`



Source: `http://richard.cyganiak.de/2007/10/lod/`

Figure 1.1: LinkedGeoData in the LOD cloud. There are now links to DBpedia and GeoNames.

daily and weekly) for download. The OSM tool *Osmosis*[8] is used to synchronize our own OSM database with the main database of OSM. Furthermore, we extended our local OSM database with additional tables and views for storing the *LGD tag mappings*. The mappings express how tags should beconverted to a lightweight ontological structure. For example, the tag (amenity, school) is mapped to (rdf:type, lgdo:School). The core of the RDF conversion is formed by *Sparqlify*[9] [3]. Sparqlify is a SPARQL-SQL rewriter and as such capable of answering SPARQL queries by rewriting them to corresponding SQL queries on the OSM replica via a set of view definitions. The *Sparqlify platform* provides additional web interfaces. Applications and mashups, such as the LinkedGeoData browser[10] can then be built against these interfaces.

1.2.1 Challenges

The challenges we face in LinkedGeoData are, to a large extent, Big Data challenges, and therefore related to

- *Large volume:* OSM's database size is about 300GB, the additional indexes used in LGD account for 1TB.
- *High velocity:* Hundreds of changes per second are made in OSM.
- *High variety:* More than 10.000 attributes and many millions of different attribute-value combinations.

[8] `http://wiki.openstreetmap.org/wiki/Osmosis`
[9] `http://sparqlify.org`
[10] `browser.linkedgeodata.org`

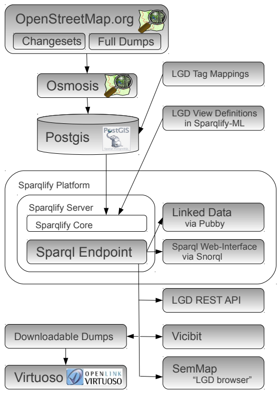


Figure 1.2: Overview of the new LinkedGeoData architecture. Most components now uniformly depend on the SPARQL endpoint provided by Sparqlify.

1.2.2 Applications and Future Perspectives

LinkedGeoData's SPARQL endpoint can be seen as the heart of the project, as this offers the possibility for anyone to freely access and query fresh RDFized OSM data. Listing 1 demonstrates a query which upon successful execution retrieves a list of helipads near hospitals. In fact, OSM data is successfully being used for humanitarian response, for instance during the *Haiti Earthquake*[11] and the *Sendai earthquake and tsunami*[12].

The advantage of LGD is, that by publishing OSM data as RDF and via SPARQL, there exists a a large set of tools for solving data integration problems. For example, SILK`http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/` and LIMES`http://aksw.org/Projects/LIMES` are state-of-the-art tools for the creation of links between knowledge bases. ANAPSID[13] and FedX[14] enable query federation over multiple SPARQL endpoints.

```
Prefix lgdo:<http://linkedgeodata.org/ontology/>
Prefix geom:<<http://geovocab.org/geometry#>
SELECT * WHERE {
    ?p a lgdo:Helipad .
    ?p geom:geometry ?pg .
    ?pg ogc:asWKT ?pgv .

    ?h a lgdo:Hospital.
    ?h geom:geometry ?hg .
    ?hg ogc:asWKT ?hgv .

    Filter(bif:st_intersects(?pgv, ?hgv, 0.5))
```

[11] `http://wiki.openstreetmap.org/wiki/WikiProject_Haiti`
[12] `http://wiki.openstreetmap.org/wiki/2011_Sendai_earthquake_and_tsunami`
[13] http://maribelacosta.com/anapsid/
[14] `http://www.fluidops.com/fedx/`

```
}
LIMIT 10
```

Listing 1: A SPARQL query asking for helipads within 500 meters from hospitals

Bibliography

[1] Auer, S., Lehmann, J., Hellmann, S.: LinkedGeoData - adding a spatial dimension to the web of data. In: Proc. of 8th International Semantic Web Conference (ISWC) (2009). DOI doi:10.1007/978-3-642-04930-9_46. URL http://www.informatik.uni-leipzig.de/~auer/publication/linkedgeodata.pdf

[2] Stadler, C., Lehmann, J., Höffner, K., Auer, S.: Linkedgeodata: A core for a web of spatial open data. Semantic Web Journal (2011). URL http://www.semantic-web-journal.net/sites/default/files/swj173_2.pdf

[3] Stadler, C., Unbehauen, J., Lehmann, J., Auer, S.: Connecting crowd-sourced spatial information to the data web with sparqlify. VLDB Special Issue on Structured, Social and Crowd-sourced Data on the Web (under review) (2012). URL http://www.semantic-web-journal.net/sites/default/files/swj173_2.pdf