

Trentino Government Linked Open GeoData: First Results

Pavel Shvaiko¹, Feroz Farazi², Daniela Ferrari³, Giuliana Ucelli³, Lorenzino Vaccari⁴,
Vincenzo Maltese², Veronica Rizzi², A. Ivanyukovich⁵

¹ *TasLab, Informatica Trentina S.p.A., Trento, Italy*

² *DISI, University of Trento, Trento, Italy*

³ *Segreteria SIAT, Autonomous Province of Trento, Italy*

⁴ *European Commission, Joint Research Center, Italy*

⁵ *Trient Consulting Group S.r.l., Trento, Italy*

Abstract. We report our work on publishing government linked open geo metadata and data of the Italian Trentino region following the Open Government Data (OGD) paradigm.

Introduction. Within the semantic geo-catalogue application [6, 7], the Autonomous Province of Trento (PAT) has published some of its core geodata accompanied with the corresponding metadata as linked open data (LOD), namely in RDF. Even though the publication of LOD is not required by the INSPIRE directive, our results can be considered as a novel good practice to this end. In fact, in parallel with the standardization and regulation effort, the implementation of INSPIRE should take into account the linked data principles, since these facilitate data harmonization. For instance, the issue is to identify the most relevant vocabularies for RDF representation of the INSPIRE metadata elements. Also geodata, modeled as INSPIRE themes, can be represented as RDF triples in order to facilitate its discovery and future re-use. Within the European Commission, the process has already started, for example, for the INSPIRE data theme “addresses” specification, which was used as a basis to model the “Address” class of the Core Location Vocabulary of the Interoperability Solutions for European Public Administration (ISA) program (<http://tinyurl.com/72538jm>).

In turn, the OGD paradigm encourages governments to publish their data in an open (from both technical and legal perspectives) manner in order to foster transparency and economic growth (through data re-use), see for example, the respective US (<http://www.data.gov>) and UK (<http://data.gov.uk/>) initiatives. In particular, the UK government published its data using open standards, e.g., RDF for representation, SPARQL endpoint for exposing (<http://www.w3.org/TR/rdf-sparql-query/>) and Dublin Core (DC) vocabulary for annotation. Essentially, the use of a SPARQL endpoint for exposing data allows for programmatic access to it and for semantic search engines (e.g., Sindice: <http://www.sindice.com/>) to discover, crawl and index it, what helps increasing its visibility and re-use. In this context, geospatial information was also published by Ordnance Survey [3]. Finally, a similar initiative was launched, among other countries, by Spain which published its GeoLinked Data [4] starting from the data sources of the Spanish National Geographic Institute (<http://www.ign.es/>).

Trentino first results. Under the recent regional deliberation n. 195/2012, PAT formally decided to adopt Creative Common Zero (public domain) license to release 161 of its geographical core datasets, such as bicycle tracks, administrative boundaries, ski areas, and CORINE land cover. Both metadata and data of these were selectively (by using fragments that facilitate its discovery and re-use) represented in a single RDF file per dataset.

Geographic metadata, originally provided following the ISO19115 standard, was encoded using the DC and DCMI-BOX (<http://dublincore.org/documents/dcmi-box/>) standard vocabularies, see an example below:

```
<rdf:Description rdf:about="http://www.territorio.provincia.tn.it/geodati/p_tn:piste_ciclabili">
  <dc:language>it</dc:language>
  <dc:mbox:westlimit>10.41</dc:mbox:westlimit>
  <dc:identifier>http://www.naturambiente.provincia.tn.it/</dc:identifier>
  <dc:format>shp</dc:format>
  <dc:rights>Dato pubblico</dc:rights>
  <dc:title>Piste ciclabili</dc:title>
```

```
<dc:creator>Dipartimento Risorse Forestali e Montane</dc:creator>
</rdf:Description>
```

Geodata from shapefiles was selectively published in RDF as well, see an example below. To express the features geographic position, the UTM coordinate system was preserved, while new terms were created only in case no suitable candidate was available in the standard vocabularies [5]. Specifically, we created the *length*, *area*, *perimeter* and *polyline* terms, which need to be provided as a resolvable URI. When available, we specified the length of the features modeled as polylines and the area and perimeter of the features modeled as polygons. Moreover, we linked the RDF (source) to the most relevant hub datasets (target), such as DBPedia and Freebase. This was done through the OWL *owl:sameAs* association. To ensure the accuracy, the links between the source and the target resources were established manually.

```
<rdf:Description rdf:about="http://www.territorio.provincia.tn.it/geodati/resource/piste_ciclabili">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  <owl:sameAs rdf:resource="http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000428308"/>
</rdf:Description>
```

Feature class piste_ciclabili (in English bicycle_tracks) and a sameAs link to the external dataset Freebase

```
<rdf:Description rdf:about="http://www.territorio.provincia.tn.it/geodati/resource/piste_ciclabili/529">
  <geontology:length rdf:datatype="http://www.w3.org/2001/XMLSchema#double">1445.8484810675</geontology:length>
  <rdfs:label xml:lang="it">Mori - torbole</rdfs:label>
  <rdf:type rdf:resource="http://www.territorio.provincia.tn.it/geodati/resource/piste_ciclabili"/>
  <rdfs:label xml:lang="it">529</rdfs:label>
  <geo:geometry rdf:resource="http://www.territorio.provincia.tn.it/geodati/resource/piste_ciclabili_529"/>
</rdf:Description>
```

Mori-Torbole, an instance of the class bicycle_tracks

```
<rdf:Description rdf:about="http://www.territorio.provincia.tn.it/geodati/resource/piste_ciclabili_529">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  <geontology:polyline>646339.346896746,5082179.74045936
  ...
  645575.851739799,5081173.68539361
  </geontology:polyline>
</rdf:Description>
```

Geometric shape (polyline) of the Mori-Torbole track

We also developed a mashup application available at <http://sgc.disi.unitn.it:8080/sgcmashup/> to observe the usefulness of the published geodata in linking and accessing different datasets. To combine information from different RDF sources, we used the DERI pipes tool [1]. The development of this mashup on top of the linked geodata took a short time (about 4 days) compared to the one required if we were to develop the same mashup without using semantic technologies.

Conclusions. We outlined our work on releasing some of the Trentino geo-data and metadata as part of the linked data. RDF was used for representing fragments of both geo-data and the respective metadata. We used well-known standards and specifications including Dublin Core for metadata, WGS84 for data and OWL for linking data to the external resources, such as DBPedia and Freebase. Some of the future work includes the definition of geometrical terms through the NewGeo Geometry vocabulary (<http://geovocab.org/geometry.html>), the addition of RDF VoID descriptions for the datasets and their publication on the LOD cloud.

Acknowledgments. This work has been supported by the Autonomous Province of Trento.

References

1. D. Le-Phuoc, A. Polleres, M. Hauswirth, G. Tummarello, C. Morbidoni: Rapid prototyping of semantic mash-ups through semantic web pipes. In *Proc. of WWW*, 2009.
2. INSPIRE Technical Architecture – Overview, 2007.
3. J. Goodwin, C. Dolbear, G. Hart. Geographical linked data: the administrative geography of Great Britain on the semantic web. *Transaction in GIS* 12(1), 2009.
4. L. M. V. Blázquez, B. Villazón-Terrazas, V. Saquicela, A. de León, Ó. Corcho, A. Gómez-Pérez. GeoLinked data and INSPIRE through an application case. In *Proc. of GIS*, 2010.
5. T. Heath, C. Bizer. Linked Data: Evolving the web into a global data space. *Morgan and Claypool*, 2011.
6. F. Farazi, V. Maltese, F. Giunchiglia, A. Ivanyukovich. A faceted ontology for a semantic geo-catalogue. In *Proc. of ESWC*, 2011.
7. P. Shvaiko, A. Ivanyukovich, L. Vaccari, V. Maltese, F. Farazi. A semantic geo-catalogue implementation for a regional SDI. In *Proc. of INSPIRE*, 2010.