Department of
Information Engineering
and Computer Science
DISI

UNIVERSITY
OF TRENTO - Italy

DISI - Via Sommarive 14 - 38123 Povo - Trento (Italy)
http://www.disi.unitn.it

TRENTINO GOVERNMENT LINKED OPEN GEODATA: FIRST RESULTS

Pavel Shvaiko, Feroz Farazi, Lorenzino Vaccari, Vincenzo Maltese, Veronica Rizzi

March 2012

Technical Report # DISI-12-012

A short version of this technical report has been published in the proceedings of the INSPIRE Conference 2012.

Trentino government linked open geodata: first results

Pavel Shvaiko[1], Feroz Farazi[2], Daniela Ferrari[3], Giuliana Ucelli[3], Lorenzino Vaccari[4], Vincenzo Maltese[2], Veronica Rizzi[2], A. Ivanyukovich[5]

[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: In this paper we report our work on publishing linked data of the Trentino Region in Italy following the Open Government Data (OGD) Initiative and in compliance with INSPIRE directive.

Introduction: As part of the Semantic Geo-Catalogue (SGC) project [15, 16], the Autonomous Province of Trento (PAT) published its geographical data as linked data in RDF. Within the INSPIRE Directive[1], it is required to have, among others, a resolvable unique identifier for each spatial object [3]. In the Linked Data [4], objects are named using resolvable HTTP URIs [5]. Therefore, the notion of Linked Data is in line with INSPIRE. In this paper we describe the methodology and design decisions we took to produce the linked data.

Open Government Data: The OGD movement encourages governments to publish their data and content under a suitable license so that others can use, reuse and redistribute it through proper attribution. The Open Knowledge Foundation (OKF) community provides a list[2] of such licenses. They are divided in two families: Open Data Commons[3] (ODC) for data and Creative Commons[4] (CC) for content. CC-Zero supports both data and content sharing and provides maximum flexibility in their use.

Linked Data: In UK the INSPIRE Directive is followed. In particular, the UK government has decided to publish their data using open standards, e.g., RDF for representation [6], SPARQL Endpoint for exposing [4, 7], DCMI (Dublin Core Metadata Initiative) vocabulary for annotation [8] and GML (Geography Markup Language) for representing geographic features [9]. Essentially, the use of a SPARQL Endpoint for exposing data allows the Semantic Web search engines (for instance Sindice[5], Swoogle[6] and Watson[7]) to discover, crawl and index the RDF data which in turn helps increasing the visibility of the data itself. Ordnance Survey[8], the national mapping

[1] http://www.ec-gis.org/inspire/
[2] http://opendefinition.org/licenses
[3] http://www.opendatacommons.org/
[4] http://creativecommons.org/
[5] http://www.sindice.com/
[6] http://swoogle.umbc.edu/
[7] http://watson.kmi.open.ac.uk/
[8] http://www.ordnancesurvey.co.uk

agency in the UK, spearheaded the publishing of geospatial information as part of the Linked Data [10]. There are similar initiatives in EU. For example, in Spain the Geo-Linked Data [12] initiative at the University Politecnica de Madrid has contributed bringing Spanish geographic and statistical information to the Linked Data. They have dealt with the data sources owned by the Spanish National Geographic Institute (IGN-E)[9] and Spanish National Statistical Institute (INE)[10]. Their dataset is linked to GeoNames[11] and DBPedia[12]. For the representation of the statistical (e.g., unemployment rate), geometrical (e.g., shape) and geo-positioning (e.g., geographical coordinates) information, Statistical Core Vocabulary (SCOVO)[13], GML and WGS84 vocabularies were used, respectively.

Trentino Linked Data: Under the regional deliberation n. 195/2012 the PAT formally decided to adopt CC-Zero to release 161 of its geographical datasets, such as streams and bicycle tracks. Data was available as shape files, while metadata (aligned with the ISO 19115 Core metadata elements) was available as XML files. To make them part of the Linked Data cloud [13], we converted both of them in RDF.

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dcmibox="http://dublincore.org/documents/dcmi-box/"
    xmlns:dc="http://purl.org/dc/elements/1.1/" >

   <rdf:Description
rdf:about="http://www.territorio.provincia.tn.it/geodati/p_tn:piste_ciclabili">
    <dc:language>it</dc:language>
    <dcmibox:westlimit>10.41</dcmibox:westlimit>
    <dcmibox:eastlimit>11.97</dcmibox:eastlimit>
    <dcmibox:southlimit>45.60</dcmibox:southlimit>
    <dcmibox:northlimit>46.60</dcmibox:northlimit>
    <dc:identifier>http://www.naturambiente.provincia.tn.it/</dc:identifier>
    <dc:format>shp</dc:format>
    <dc:rights>Uso limitazione: nessuna limitazione. Altri vincoli: Dato pubblico</dc:rights>
    <dc:title>Piste ciclabili</dc:title>
    <dc:creator>Dipartimento Risorse Forestali e Montane</dc:creator>
    <dc:version>1.0</dc:version>
    <dc:date>2008-09-26</dc:date>
  </rdf:Description>
</rdf:RDF>
```

Fig. 1. RDF representation of the Metadata

As shown in Fig 1 metadata has been encoded using the DCMI and DCMI-BOX[14] standard vocabularies. We focused on those which fall in the intersection of INSPIRE/ISO Core metadata and Dublin Core. They were grouped under a resource, which was given a URI generated by appending the *file identifier* (`p_tn:piste_ciclabili`) metadata attribute to the namespace URI (`http://www.territorio.provincia.tn.it/geodati/`) for the Trentino datasets. The metadata resource language, online locator, distribution format, title, responsible organization, version and creation date were mapped to dc:language, dc:identifier, dc:format, dc:rights, dc:title, dc:creator, dc:version and

[9] http://www.ign.es/
[10] http://www.ine.es/
[11] http://www.geonames.org/
[12] http://dbpedia.org/
[13] http://vocab.deri.ie/scovo/
[14] http://dublincore.org/documents/dcmi-box/

dc:date, respectively; the geographic bounding box attributes west bound longitude, east bound longitude, south bound latitude and north bound latitude were mapped to dcmibox:westlimit, dcmibox:eastlimit, dcmibox:southlimit and dcmibox:northlimit, respectively.

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:geontology="http://www.territorio.provincia.tn.it/geodati/ontology/"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#" >

  <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>

  <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>

  <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
                         646329.929020191,5082161.84683082
                         ...
                         645576.090351533,5081173.94569307
                         645575.851739799,5081173.68539361
    </geontology:polyline>
  </rdf:Description>
</rdf:RDF>
```

Fig. 2. RDF representation of the data

As shown in Fig. 2 in the data encoding we used WGS84 for geographically located objects. At the time of publishing, new terms were created only in the cases of their unavailability in the standard vocabularies (see [14]). Basically, we created *length*, *area*, *perimeter*, *polyline*, *polygon* and *geometry*, which need to be provided a resolvable URI. When available, with polylines we attached length and with polygons we attached area and perimeter. Following best practices and standards [1], we linked the RDF (source) to the most relevant vocabularies (target), i.e. DBPedia and Freebase[15] and for this purpose *owl:sameAs* links were used. The mapping between the source and target resources was accomplished manually to obtain human level accuracy. RDF data was modelled as follows:

- A namespace called geontology was created with the following URI:
  geontology="http://www.territorio.provincia.tn.it/geodati/ontology/"
  Note that this URI should be made resolvable.
- Each RDF file consists of one class and a set of instances.

[15] http://www.freebase.com/

Note that RDF provides the flexibility to treat instances as classes.

- Well-established vocabularies were referred (e.g., WGS84) for the existing terms and for the newly defined related terms.
- The RDF data model was produced using Jena[16] Semantic Web tool.

We developed a mashup application to prove the usefulness and effectiveness of the published data and of the Linked Data in general in linking and accessing to different datasets. The purpose of this application is to support the following scenario:

Robert is in a summer trip to Trento cycling in the bicycle path between Trento and Riva del Garda. Once he arrived in the lakefront region of the Mori-Torbole bicycle track, he is fascinated by the splendid natural beauty of the lake and the panoramic beauty of the mountains, which made him interested to know more about the panoramic views of the other parts of the bicycle track and the nearby hotels to stay there for some days. Cycling in the summer noon made him thirsty, hence he is eager to know the location of the drinking water fountains in the vicinity of the bicycle track.

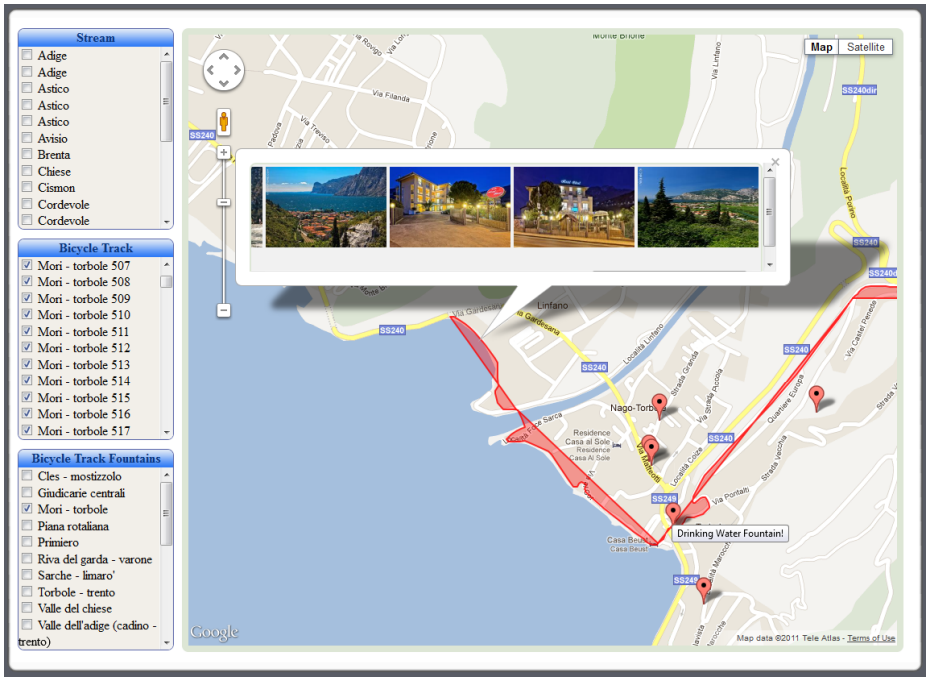

Fig. 3. The mashup developed to support the cyclist-tourist scenario

Fig 3 provides a snapshot of the mashup application supporting this scenario. Streams (e.g., Adige), bicycle tracks (e.g., Mori - Torbole 507) and bicycle track

[16] http://incubator.apache.org/jena/

fountains are shown on the left as a list of check boxes, where the numbers to the right of the tracks represent the identifiers of the track parts which constitute the whole track. Selected streams, bicycle tracks and fountains are displayed using Google Maps[17] as polygons, polylines and markers, respectively. By clicking on a bicycle track it is possible to visualize a set of images of the nearby hotels and panoramic views. We collected images from Flickr[18] and we gathered information about drinking water fountains from Open Street Map through LinkedGeoData[19]. For gathering, combining and merging information from different sources published in RDF, we used DERI Pipes[20], a state-of-the-art Semantic Web mashup tool [2]. The development of this mashup on top of the 5-star data took only 4 days, while developing a similar mashup without semantic technologies would take months.

Conclusion: We briefly presented our work on bringing Trentino OGD as part of the Linked Data. In this work, RDF was used for representing data and metadata. Among the vocabularies we used Dublin Core for representing the metadata, WGS84 for representing data and OWL to link data to the external sources DBPedia and FreeBase. New terms were defined only when they were not available in existing vocabularies. Nevertheless, for geometrical terms NewGeo Geometry[21] can be referred in the future. Our future plan includes, among others, adding RDF VoID descriptions [17] of the datasets and putting things in the lod cloud[22].

References

1. C. Bizer, R. Cyganiak, T. Heath. How to publish Linked Data on the Web. http://www4.wiwiss.fuberlin.de/bizer/pub/LinkedDataTutorial/.
2. D. Le-Phuoc, A. Polleres, M. Hauswirth, G. Tummarello, C. Morbidoni: Rapid Prototyping of Semantic Mash-ups through Semantic Web Pipes. WWW, 2009.
3. INSPIRE Technical Architecture – Overview, 2007.
4. T. Berners-Lee. Linked Data. http://www.w3.org/DesignIssues/ LinkedData.html, 2006.
5. J. Sheridan, J. Tennison. Linking UK Government Data. WWW2010 workshop: Linked Data on the Web (LDOW), ACM, 2010.
6. D. Brickley, R. V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation, http://www.w3.org/TR/rdf-schema, 2004.
7. SPARQL Query Language for RDF. K. G. Clark, L. Feigenbaum, E. Torres, Editors, W3C Recommendation, http://www.w3.org/TR/rdf-sparql-query/, 2008.
8. Dublin Core Metadata Element Set, Version 1.1: Reference Description, http://dublincore.org-/documents/dces/, 2010.
9. S. Cox, P. Daisey, R. Lake, C. Portele, A. Whiteside. OpenGIS Geography Markup Language (GML) Implementation Specification, version. 3.1.0, 2004.

[17] http://maps.google.com/
[18] http://www.flickr.com/
[19] http://linkedgeodata.org/
[20] http://pipes.deri.org/
[21] http://geovocab.org/geometry.html
[22] http://linkeddata.org/

10. 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.
11. F. Lopez-Pellicer, M. Silva, M. Chaves, F. Javier Zarazaga-Soria, P. Muro-Medrano, Geo linked data. Proceedings of the 21st International Conference on Database and expert systems applications DEXA'10, Vol. 6261, Springer-Verlag, pp. 495-502, 2010.
12. 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 GIS, pp. 446–449, 2010.
13. C. Bizer, T. Heath, T. Berners-Lee. Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, Special Issue on Linked Data, 5(3): 1-22, 2009.
14. T. Heath, C. Bizer. Linked Data: Evolving the Web into a Global Data Space. http://linkeddatabook.com/.
15. Farazi, F., Maltese, V., Giunchiglia, F., Ivanyukovich, A. A faceted ontology for a semantic geo-catalogue. 8th Extended Semantic Web Conference (ESWC), 2011.
16. Shvaiko, P., Ivanyukovich, A., Vaccari, L., Maltese, V., Farazi, F. A semantic geo-catalogue implementation for a regional SDI. INSPIRE conference, 2010.
17. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J. Describing linked datasets with the VoID vocabulary, 2011.