Geospatial Semantics for Topographic Data

E. Lynn Usery

U.S. Geological Survey

Abstract. Whereas *The National Map* of the U.S. Geological Survey is based on data models and processes of geographic information systems, there is a current effort to explore the potential of semantically-based geospatial data using the Resource Description Framework (RDF) triple model of the Semantic Web. Advantages of the RDF approach include the ability to encode richer semantics, such as part-whole relations within features and geometric, topologic, thematic, and temporal relations between features. Procedures for the automatic conversion of vector data to RDF have been developed. Raster datasets also can be converted but require human interaction to define geographic features and their associated characteristics, which are then converted to RDF. Geospatial data in RDF can be accessed, queried, analyzed, and mapped based on the features, characteristics, and geometry contained in the triple model.

Keywords: Geospatial semantics, RDF, geographic features

1. Introduction

The U.S. Geological Survey (USGS) began development of *The National Map* in 2001 (Kelmelis *et al.*, 2003; USGS 2013a). It was initially conceived as a repository of geospatial data in seamless nationwide databases in geographic information system (GIS) data models for eight data layers: transportation, hydrography, boundaries, structures, geographic names, land cover, elevation, and orthoimagery. All data would be in the public domain and accessible from a map viewer interface with supported Web services and download capabilities. The databases were to be current and provide the basis for a new generation of

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U. S. Government.

topographic maps. From 2001 to 2008, the USGS developed the data for the eight data layers from existing sources, augmented the data with new collection, and developed seamless, integrated nationwide databases, making them available in the public domain. In 2009, the USGS began delivery of US Topo, the new topographic map series for the United States, automatically generated from the databases of *The National Map* (USGS, 2013b).

Simultaneous to the development of *The National Map*, the World-Wide Web Consortium instantiated the concept and working environment of the Semantic Web, which was also launched in 2001. In the years since these initial developments, both *The National Map* and the Semantic Web have made significant advances and there has been considerable research to adapt geospatial data to the Semantic Web. Within the USGS, advances have been made in direct conversion of GIS databases from *The National Map* to the Resource Description Framework (RDF) triple model commonly used on the Semantic Web.

The purpose of this paper is to present an approach adopted by the USGS for conversion of GIS databases of *The National Map* to the RDF format of the Semantic Web. The next section of the paper describes the conversion of vector GIS data in relational databases to RDF format. The third section provides a synopsis of raster conversion approaches based on hand coding of attributes and relationships. The fourth section shows how geometry is handled in RDF and provides some examples of queries and mapping geographic features from the RDF triplestores. A final section draws some conclusions and gives some future possibilities for geospatial data in RDF.

2. Vector Data Conversion to RDF

2.1. Vector-based GIS datasets are composed of objects either defined as point, line, or areas or as actual geographic entities, such as roads and streams. Attributes and relationships, particularly topology, are commonly stored in relational database tables, which can easily be converted to the subject, predicate, object of the triple model of the Semantic Web. An automatic conversion process is possible in which the rows of the table become subjects, the columns become predicates, and the cell values become the objects. The USGS has implemented such an approach and has made data for hydrography, transportation, boundaries, and structures from nine watersheds available in this form for specific research test sites in the United

States (Varanka *et al.*, 2011; Usery and Varanka, 2012). Geographic names have been converted for entire country. Access is through a project web site, http://cegis.usgs.gov/ontology.html, and through a SPARQL Protocol and RDF Query Language (SPARQL) Endpoint http://usgs-ybother.srv.mst.edu:8890/parliament. Further, a conversion program has been developed and made available that performs this conversion for any specified area of *The National Map* databases for vector datasets including hydrography and transportation. USGS has developed an online, publically accessible tool to convert data from the relational databases of *The National Map* to RDF triple form. The user simply specifies the area to be converted by either a named reference, Polk County, Missouri, USA, for example, or from a polygon boundary in shapefile or Well Known Text (WKT) format.

3. Raster Data Conversion to RDF

Raster data poses a more significant challenge since GIS data in this format commonly use a field view and do not identify specific geographic entities that can be encoded as features. Initial work to convert raster data to RDF and capture semantic relationships has used an approach of examining named geomorphic features and hand coding the relationships between features while maintaining a minimum bounding rectangle as the geometric footprint of the features in the raster datasets, such as terrain elevation and orthographic images. For example Figure 1 shows the geometric footprint of Last Chance Bench, a terrain feature, with the footprint (minimum bounding rectangle) extracted from a USGS 7.5 minute topographic map from US Topo. Table 1 shows the attributes and relationships associated with the bench. Note that there is no boundary that defines the feature. Its extent is determinable only by the shape of the contours, the image background, and the placement of the name. Once identified coding the feature, attributes, and relationships in RDF is a simple matter. To simplify the presentation in the table, numbers have been used for the stream identifiers, whereas in an actual implementation as in Figure 2, a Uniform Resource Identifier (URI) for each stream is used. Whereas this approach allows features to be identified and coded in RDF, it is laborious and time intensive. It would be a near impossible task to code all terrain features in this manner, especially since most terrain features are not named. Other approaches are being examined including developing a formal terrain ontology using concepts from surface theory and geomorphology (SOCoP, 2012).

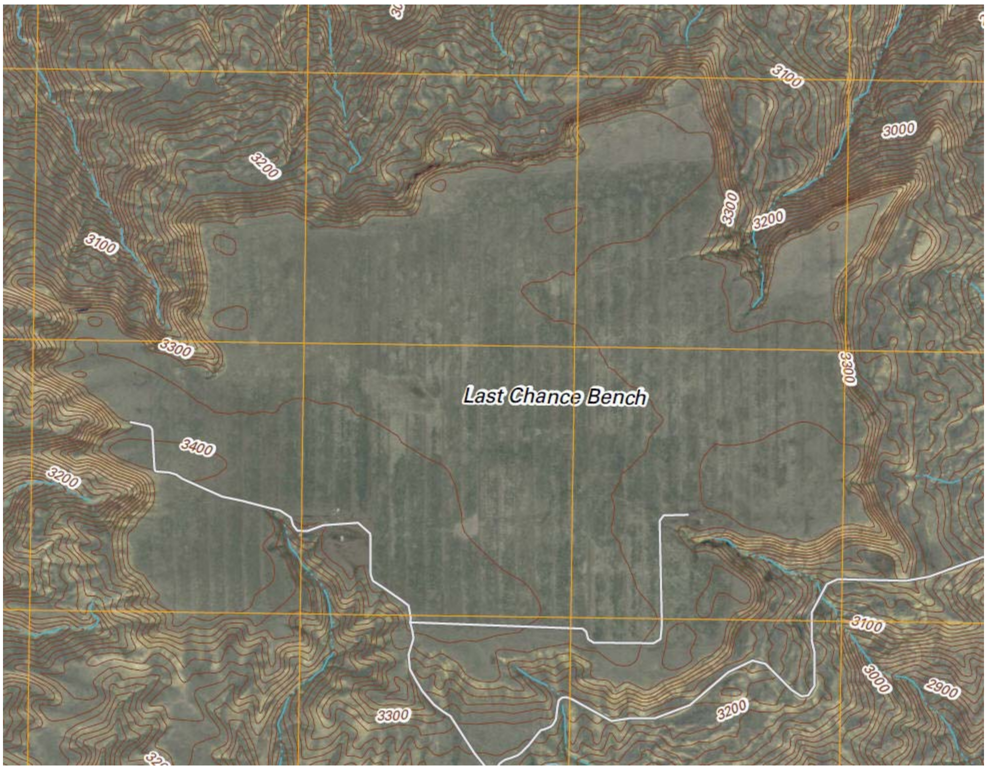

Figure 1. A geomorphic feature, Last Chance Bench, represented on a topographic map only by a name. In a raster digital elevation model, not even the name is included.

Feature Instance	Attributes	Relationships
Last Chance Bench	Elevation 3400 ft	Adjacent road
		Head of streams: 1,2,3,4,5,6,7,8,9,10,11

Table 1. A raster geomorphic feature, Last Chance Bench, with its attributes and relationships.

```
@prefix ogc: <http://www.opengis.net/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix geoname: <http://www.geonames.org/ontology#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dcterms: <http://purl.org/dc/terms/>
@prefix dbpedia: <http://dbpedia.org/ontology/>
@prefix geo: <http://www.opengis.net/ont/OGC-GeoSPARQL/1.0/>
@prefix usgsTopo: <http://cegis.usgs.gov/TopoVocab/1.0/Terrain#> .
@prefix usgs: <http://cegis.usgs.gov/ontology/instances#> .

<http://cegis.usgs.gov/ontology/instances> a owl:ontology

usgs:_773239    a usgsTopo:bench;
    a geo:Feature ;
    geo:hasGeometry usgs:_773239geo ;
    geoname:name Last Chance Bench"
    rdfs:comment "A topographic bench";
    dcterms:identifier "773239" ;
    dcterms:description " An area of relatively level land on the flank of an elevation
such as a hill, ridge, or mountain where the slope of the land rises on one side and
descends on the opposite side (level)"

usgs:_773239geo a geo:Geometry ;
    usgsTopo:hasUTM "12 581864 5286804";
    usgsTopo:hasUSNG "12T WT 81874 86808  (NAD 83)";
    usgsTopo:hasMBR "Max E 583000m Min E 580000m Max N 5288200m Min
N 5285360m";
    dbpedia:MaximumElevation "3400ft";

usgs:_tigerA487336 a usgsTopo:Road;
  hasGeometry usgs:_tigerA487336geo;
  geoname:nearby usgs:_773239

usgs:_tigerA487336geo a geo:Geometry;

usgs:_78950327 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.

usgs:_78950415 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.

usgs:_78950591 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.

usgs:_78950729 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.
```

```
usgs:_78950725 a usgsTopo:Canal;
          rdfs:label "Last Chance Canal";
     geoname:nearby usgs:_773239 .

usgs:_78950683 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.

usgs:_78950715 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.

usgs:_78950891 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.

usgs:_78950565 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.

usgs:_78950591 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.

usgs:_78950433 a usgsTopo:Stream;
          geoname:nearby usgs:_773239.
```

Figure 1. An example RDF coding of Last Chance Bench based on the attributes and relationships from Table 1.

4. Geometry in RDF

Coding geometry in RDF for both vector and raster features relies onWKT and the Geography Markup Language (GML), the two techniques supported by the Semantic Web, SPARQL, and its extension GeoSPARQL, developed as a standard by the Open Geospatial Consortium (OGC). Using these methods it is possible to develop graphics, examine topological relationships, and perform spatial analyses on geographic data stored as RDF triples. In particular, GeoSPARQL supports geometric operations and the eight topological relationships of the Simple Features Relations Family of the OGC. The USGS has tested these protocols with the research datasets that have been converted from vector relational tables and is examining these approaches to handling geometry and topology for entities defined on raster data. Initial results demonstrate that it is possible to use WKT and GML to create maps and perform simple spatial analysis functions.

The USGS has designed a graphical interface that allows entry of the SPARQL or GeoSPARQL query with the result mapped onto an orthographic image backdrop. That interface is shown in Figure 3.

4.1 An Example Query and Map Result

As an example, combining USGS data with Environmental Protection Agency (EPA) data is presented. The search will find EPA hazardous sites within 5 km of the Pittsburg Firehouse near Sentinel, Missouri, USA. The RDF query follows with the text results in Table 2 and the mapped results in the graphical user interface in Figure 3.

First define the needed prefixes which allow use of standard namespaces on the Semantic Web:

```
PREFIX geo: <http://www.opengis.net/geosparql#>
PREFIX geof: <http://www.opengis.net/geosparql/function/>
PREFIX gml: <http://www.opengis.net/gml#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gnis: <http://cegis.usgs.gov/rdf/gnis/>
PREFIX gnisf: <http://cegis.usgs.gov/rdf/gnis/Features/>
PREFIX nhd: <http://cegis.usgs.gov/rdf/nhd/>
PREFIX nhdf: <http://cegis.usgs.gov/rdf/nhd/Features/>
PREFIX gu: <http://cegis.usgs.gov/rdf/gu/>
PREFIX guf: <http://cegis.usgs.gov/rdf/gu/Features/>
PREFIX category: <http://dbpedia.org/class/yago/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX units: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX dgtwc: http://www.data.gov/semantic/data/alpha/1050/dataset-1050.rdf#
```

The query is then entered and executed.

```
SELECT DISTINCT
?name ?wkt1
WHERE {
```

```
GRAPH <http://cegis.usgs.gov/rdf/> {
  # Match features with type EPA DataEntry
  ?feature rdf:type   <http://data-gov.tw.rpi.edu/2009/data-gov-
twc.rdf#DataEntry> .
  ?feature geo:asWKT ?wkt1 .
  ?feature dgtwc:primary_name ?name .
  # Get geometry of the firehouse
  <http://cegis.usgs.gov/rdf/struct/Features/10474482> geo:hasGeometry
?geo .
  ?geo geo:asWKT ?fire_wkt .
 # Create a 5km buffer around the firehouse
 BIND (geof:buffer(?fire_wkt, 5000, units:metre) AS ?fire_buff)

 # Restrict matches to the buffer
 FILTER(geof:sfContains(?fire_buff, ?wkt1))
}
}
```

The text result of the query is shown in Table 2 with the graphical result in Figure 3.

name	**wkt1**
ASH GROVE AGGREGATES, INC	POINT(-93.304139 37.823306))
DALE & SHELLY WHITESIDE	POINT(-93.295654 37.858091))
MDNR, DIV OF STATE PARKS	POINT(-93.300556 37.833889))

Table 2. Text results of the query for EPA hazardous sites within 5 km of the Pittsburg firehouse.

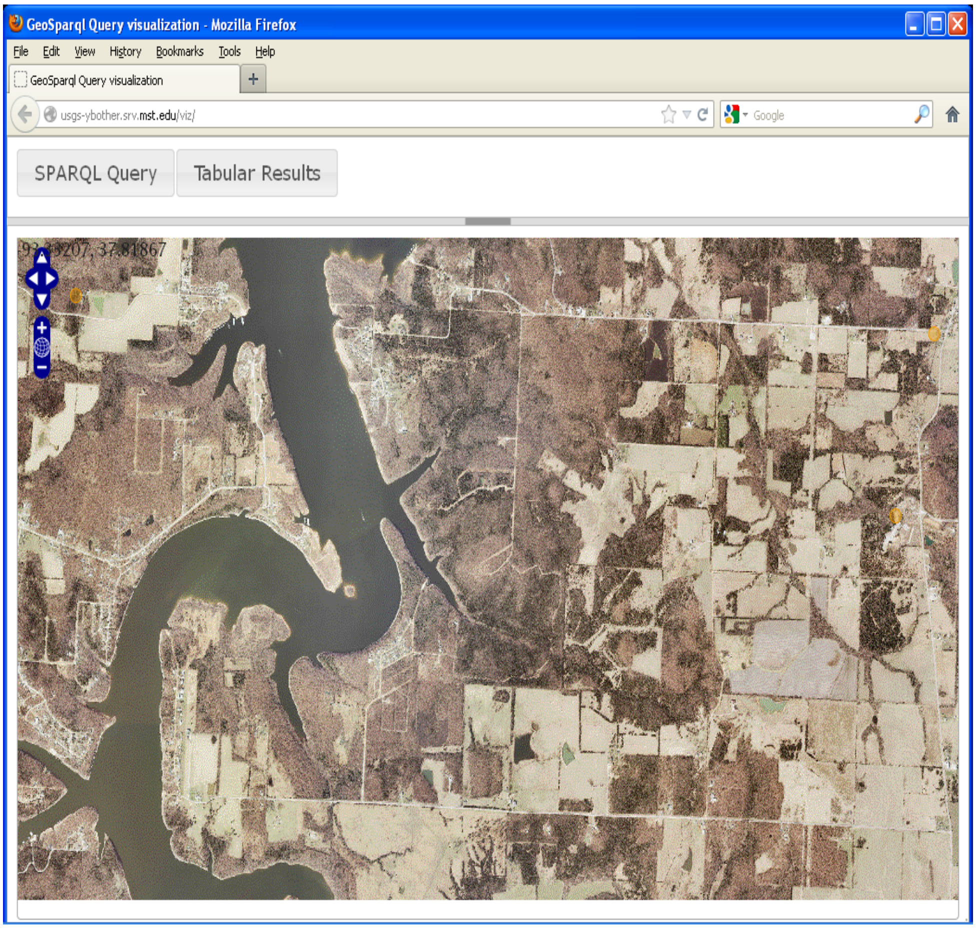

Figure 3. Graphical result of the query for EPA hazardous sites within 5 km of the Pittsburg firehouse. The EPA sites are shown as small orange circles.

5. Conclusion

The Semantic Web offers the possibility of encoding geospatial data with richer semantics and allows use of inferencing to create new data and information. Geometry is implemented in the RDF model of the Semantic Web and can be used for mapping. GeoSPARQL provides an ontology that supports geometric and topological operations, which allows creation of graphical results from queries. The RDF linked data process supports integrating data from multiple sources and organizations to create environmental and thematic maps.

References

J. A. Kelmelis, M.L. DeMulder, C. E. Ogrosky, N J. Van Driel, and B.J. Ryan, 2003. *The National Map:* From Geography to Mapping and Back Again Photogrammetric Engineering & Remote Sensing, Vol. 69, No. 10, pp. 1109–1118.

SOCoP, 2013. Spatial Ontology Community of Practice, http://www.socop.org/, accessed 4/15/2013.

Usery, E.L. and Varanka, D.E., 2012, Design and Development of Linked Data for The National Map. The Semantic Web Journal. http://www.semantic-web-journal.net/content/design-and-development-linked-data-national-map.

USGS, 2013a. *The National Map*, U.S. Geological Survey, http://nationalmap.gov/, accessed 4/15/2013.

USGS, 2013b. US Topo Quadrangles – Maps for America, U.S. Geological Survey http://nationalmap.gov/ustopo/, accessed 4/15/2013.

Varanka, D.E., Carter, J.J., Shoberg, T., and Usery, E.L., 2011, Topographic Mapping Data Semantics Through Data Conversion and Enhancement, in Sheth, Amit and Ashish, Naveen, Eds., *Geospatial Semantics and the Semantic Web, Foundations, Algorithms, and Applications*. Semantic Web and Beyond, Vol. 12: Computing for Human Experience, Springer. p. 145-162.