Design and development of linked data from *The National Map*

E. Lynn Usery[*] and Dalia Varanka
U.S. Geological Survey, 1400 Independence Road, Rolla, MO, USA

Abstract. The development of linked data on the World-Wide Web provides the opportunity for the U.S. Geological Survey (USGS) to supply its extensive volumes of geospatial data, information, and knowledge in a machine interpretable form and reach users and applications that heretofore have been unavailable. To pilot a process to take advantage of this opportunity, the USGS is developing an ontology for *The National Map* and converting selected data from nine research test areas to a Semantic Web format to support machine processing and linked data access. In a case study, the USGS has developed initial methods for legacy vector and raster formatted geometry, attributes, and spatial relationships to be accessed in a linked data environment maintaining the capability to generate graphic or image output from semantic queries. The description of an initial USGS approach to developing ontology, linked data, and initial query capability from *The National Map* databases is presented.

Keywords: Geospatial semantics, topographic data, The National Map, SPARQL endpoint, geographic features

1. Introduction

The USGS is a primary supplier of geospatial and environmental datasets that are used extensively in mapping, planning, resource and land management, emergency response, and many other applications. A sampling of these public domain data is presented in Table 1 with URLs for access. Use of these data often requires combining one or more of these datasets or combining these data with user-generated data. Since the data exist in many different formats, some proprietary, the integration or conflation of the data for use in a specific application requires significant data processing and manipulation by the user. *The National Map* (Fig. 1), which is the 21st century topographic map for the USGS, is viewed as a primary basis for these integration processes.

The Semantic Web offers an alternative approach to data formatting, access, and integration for use in applications [55]. By use of the standard triple model of the Resource Description Framework (RDF) of the Semantic Web [54], applications are able to link to other data and to use and share data effectively to answer queries and support specific applications [20]. The USGS has begun exploring the potential of the Semantic Web, particularly for geospatial data access, integration, synthesis, and use in applications. This paper provides a case study description of that initial exploration with the following three primary objectives:

– To present a USGS approach to building semantics for topographic geospatial data through the use of a taxonomy, ontology, relations (particularly spatial), and data formatting for semantic access, query, and retrieval including geometry,

---

[*] Corresponding author. E-mail: usery@usgs.gov.

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

Table 1

Sample datasets managed by the U.S. Geological Survey

| Dataset	Geometry/Format	Attribution/Scaling	URL
National Hydrography Dataset (NHD)	Vector	Discrete/nominal	http://viewer.nationalmap.gov/viewer/nhd.html?p=nhd
National Transportation Dataset	Vector; tables	Discrete/nominal	http://viewer.nationalmap.gov/viewer/ http://gisdata.usgs.net/website/MRLC/viewer.htm
National Boundaries Dataset	Vector	Discrete/nominal	http://viewer.nationalmap.gov/viewer/
National Structures Dataset	Vector	Discrete/nominal	http://viewer.nationalmap.gov/viewer/
Geographic Names Information System (GNIS)	Vector	Discrete/nominal	http://geonames.usgs.gov/domestic/download_data.htm
National Elevation Dataset (NED)	Raster	Continuous/ratio	http://viewer.nationalmap.gov/viewer/ http://seamless.usgs.gov/website/seamless/viewer.htm
National Digital Orthophotos	Raster	Continuous/ interval	http://www.ndop.gov/data.html http://viewer.nationalmap.gov/viewer/ http://gisdata.usgs.net/website/MRLC/viewer.htm
National Land Cover Dataset (NLCD)	Raster	Discrete/nominal	http://viewer.nationalmap.gov/viewer/ http://gisdata.usgs.net/website/MRLC/viewer.htm
Global Land Cover Dataset	Raster	Discrete/nominal	http://landcover.usgs.gov/landcoverdata.php
LiDAR	Point	Continuous/ratio	http://viewer.nationalmap.gov/viewer/
Satellite images	Raster	Continuous/ interval	http://edcsns17.cr.usgs.gov/NewEarthExplorer/ http://glovis.usgs.gov/
Hazards (Earthquakes, Volcanoes)	Graphics	Multiple forms	http://earthquake.usgs.gov/hazards/ http://volcanoes.usgs.gov/activity/status.php
Minerals	Vector; text	Discrete/nominal	http://mrdata.usgs.gov/; http://tin.er.usgs.gov/mrds/ http://tin.er.usgs.gov/geochem/ http://crustal.usgs.gov/geophysics/index.html
Energy	Vector; graphics databases;	Multiple forms	http://energy.usgs.gov/search.html
Landscapes and Coasts	Reports	Discrete/nominal	http://geochange.er.usgs.gov/info/holdings.html
Astrogeology	Databases	Discrete/nominal	http://astrogeology.usgs.gov/DataAndInformation/
Geologic Map Database	Vector; maps; text	Discrete/nominal	http://ngmdb.usgs.gov/
Geologic Data Digital Data Series	Maps; tables	Discrete/nominal	http://pubs.usgs.gov/dds/dds-060/
National Water Information System	Graphics; charts; tables	Continuous/ratio	http://wdr.water.usgs.gov/nwisgmap/
Floods and High Flow	Graphics; charts; tables	Continuous/ratio	http://waterwatch.usgs.gov/new/index.php?id=ww
Drought	Graphics; tables	Continuous/ratio	http://waterwatch.usgs.gov/new/index.php?id=ww
Monthly Stream Flow	Graphics; tables	Continuous/ratio	http://waterwatch.usgs.gov/new/index.php?id=ww
Ground Water	Vector; tables; graphics;	Continuous/ratio	http://waterdata.usgs.gov/nwis/gw http://groundwaterwatch.usgs.gov/
Water Quality	Graphics	Continuous/ratio	http://waterdata.usgs.gov/nwis/qw http://waterwatch.usgs.gov/wqwatch/
National Biological Information Infra-structure (NBII)	Graphics; vector; geodatabases	Multiple forms	http://www.nbii.gov/portal/server.pt/community/nbii_home/236
Vegetation Characterization	Vector; text; graphics; databases; photos;	Multiple forms	http://biology.usgs.gov/npsveg/
Wildlife	Vector; text; graphics; images; video	Multiple forms	http://www.nwhc.usgs.gov/
Invasive Species	Vector; reports; databases; graphics, image	Multiple forms	http://www.nbii.gov/portal/server.pt/community/invasive_species/221
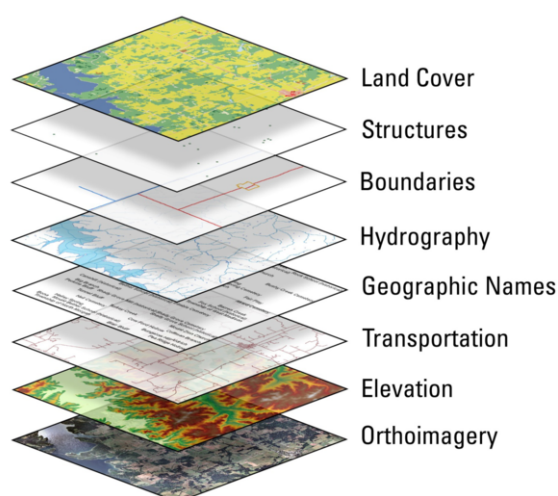
Fig. 1. Nationwide data-layers of *The National Map.*

– To show an initial conversion of data to RDF to provide interaction with the potential semantic user community, and

– To provide an approach for connecting semantics with the geometry of both vector objects and raster pixels that allows generation of graphic output in the form of maps or images as the result of queries based on semantics.

The remainder of the paper is organized as follows. Section 2 provides an overview of previous research focused on conversion of topographic data to the Semantic Web. Section 3 introduces the ontology for *The National Map* and describes the general approach to building semantics for USGS geospatial data. Section 4 describes an initial conversion of geospatial data for point and vector objects to RDF and an approach for raster data conversion to RDF. Section 5 describes a process for connecting the semantics and geometry and provides a method to access, download, and query the converted data with SPARQL Protocol and RDF Query Language (SPARQL) with a sample result. Section 6 presents conclusions based on the current work and directions for future research.

2. Previous research

A sample of the anticipated problems to be addressed by a USGS semantic approach is rooted in the broader geographic information science research agenda [52]. Specific solutions to challenges of establishing spatial semantics, designing ontology, and converting existing and new data sources to triples build on research findings reported in geospatial, ontological, and semantic literature. Examples of existing research in these areas are briefly documented below.

Topographic data are a subset of geospatial data. The national mapping agency of Great Britain, the Ordnance Survey (OS), has published ontologies and a number of research papers on various aspects of relating topography and geography to geospatial semantic technology. Some of these topics are the extraction of RDF data and OWL files from relational databases, conceptual ontology, and reasoning software [11,30]. In the context of science-driven national mapping agencies, similar to the USGS, Broderic [3] developed a framework for geographical categorization that integrates the range of topographical feature categories with the foundational, upper-level ontology DOLCE [26] and aligned with the OntoClean analytical method [17,53] see also [21]. Semantic richness is created by category criteria based on such characteristics as feature qualities, processes, roles, and relations.

An important approach in ontology design stems from temporal, activity, or event-based geographical representation [38]. These ontologies are presented as aligned with geographic theory of human-environmental interactions. Ontological representations are in part based on the forces and motivations driving events and actions in space, and themselves are influenced by intentions that impact the design of the semantic information and representation [6,9]. Though these intentional aspects of ontology development are an influence on topographical semantics represented by *The National Map*, the ontology approach applied in this research is based on natural language discourse of topographical features.

Semantic interoperability is a broad field of research for purposes of linking data across a semantic network. Spatial reference systems were conceptualized to provide a framework for connecting data [18,22]. Crucial aspects of data integration require the ontology of content data characteristics, such as the data resolution affecting geographical feature detail, data sources and uncertainty, or data maintenance [10].

Technical formalizations have emerged that are centered on linked geospatial or geoinformatics data [1,29]. The GeoVocab group defined a vocabulary for geometric coordinates and spatial object relation properties [36]. Though informal, GeoVoCamps

have produced vocabulary developments for scales, complex geometries, metadata, and temporal change [19] and [31].

3. Ontology and semantics development for *The National Map*

The ontology development combines a top-down approach based on the organization of general categories taken from standard feature classes and bottom-up approaches shaped by legacy data models. Some categories, such as transportation, which is not feature-based, require more work to align the conceptual and database models than others that are feature-based, such as NHD. The vocabulary of topographic features, to be represented as triple subjects and/or objects, was developed from standard feature list sources derived from more than a century of topographic feature data collection [41,42]. The semantic commitments of these feature lists were discussed and debated with time in a centralized way within the USGS, with input from a wide range of user communities [37,40]. Feature terms were reviewed for currency and relevance to the geographical areas within the domestic United States, so that terms such as "demilitarized zone" were edited from the list. Features that have become common since the development of the standards, such as 'windfarm', were reviewed as new vocabulary without the full development and review of a new standard. Features are classified into six taxonomic modules; terrain, surface water, ecological regime, structures, divisions, and events [49]. These reflect topographic science modeling needs and closely resemble the geographic information system (GIS) thematic layers of *The National Map*. The classification was guided with regard to regional context, feature morphology as natural or engineered structures, and descriptive attributes, such as shape and texture (fluid vs. frozen), in accordance with empirical experience and scientific concepts. The digital files form a vocabulary in OWL format, and consist of feature type classes under the taxonomic module domain. Each class has a URI, a definition, the definition source from on-line documentation, and an initial logical axiom list. The hierarchy is flat [44]. The URIs will be released to the public in the near future.

The actual implementation of conceptual systems from legacy data models is complicated by the individually created data layers contributed by partners.

For example, *The National* Map includes data from the U.S. Census Bureau, a Federal partner. Thematic integration of *The National Map* data layers occurs to support graphic map production of the U.S. Topo product (http://nationalmap.gov/ustopo/). Data layers, such as the National Hydrography Dataset (NHD), closely resemble the USGS ontology because the NHD data model defines features [43,48]. Other layers, such as transportation, are poorly matched to the conceptual ontology because they were not developed under feature-based system guidelines.

The legacy semantics extracted from standards lists that were originally developed for topographic mapping and digital data are simplistic compared to the semantic richness potentially available through the geospatial semantic web [12,50]. Engineering semantic topographic data allows complexity and decomposition that was difficult to produce in layer-based systems. The representation of topography combines natural and built-up (human-constructed) features in complex assemblages. Complex features require spatial relations among their basic components, such as the relation between an airport runway and control tower, but together build the complex feature identity. Spatial relations are often considered to form the predicate between semantically distinct feature subjects and objects of triples, but topographic features and their relations together form the semantics of complex features. Complex features are particularly common in the largest group of topographic features in the USGS vocabulary, built-up structures. In these cases, the base vocabulary allows relating simple classes into complexes for ontology design patterns (ODP) [15,16]. ODP have spatial relations that are essential to feature meaning, but a greater variety of spatial relations can be applied between distinct features when ODP are reused as specific instances.

In addition to quantitative spatial relations of location, such as coordinate pairs or geometric distances between features, spatial relation terms for the ontology development are also drawn from a set of Open Geospatial Consortium (OGC) standards for topological relations, mereological models, and verb/preposition pairs identified from the topographic feature type standards [27]. Samples of USGS topographic data reside in a triple store enabling topological reasoning according to the OGC GeoSPARQL standard [32].

Topographic features may specifically include spatial relations within the scope of the feature class

meaning, although the relation term may vary. For example, a tributary is a body of water that flows into a larger stream, or in the science vocabulary, 'drains' into another stream. In such cases, the appropriate spatial relation can be modeled with mereo-topological relations, such as 'part' or as a network 'connects', or with logical concepts, such as the Web Ontology Language (OWL) Functional-Property relation [8,39]. The logical axioms to be applied to the topographic triples are the W3C standards and functionalities offered by specific reasoning software platforms.

To capture spatial relations that support semantic identity, predicates in the form of verb/preposition pairs are presently (2011) being researched [7] in which preposition semantics reflect geometric cognition. Several categories of relations were found, including descriptive terms, such as aligned, depth, sloped, or narrowing; geometric terms, e.g., angled, confluent, curved, or extend; generative (process) terms, such as eroded, forced, suspended, and

swing; and terms of intentionality, including established, determined, designated, and defined.

4. Initial data conversion approach

The USGS approach to using the Semantic Web is to convert specific datasets from *The National Map* to RDF and make these data available for download and/or direct query in the RDF format. As a pilot project, the USGS selected nine test areas based on specific geographic characteristics, extracted all data of the eight layers of *The National Map* for these areas, and converted the vector and point data to the Geography Markup Language (GML) based on the OGC standard [33]. The nine research test areas include six watershed sub-basin areas defined from the NHD that reflect differing combinations of physiography and climate (Fig. 2). In addition to the watershed areas, the sites include three urban areas, Atlanta, Georgia; St. Louis, Mis-
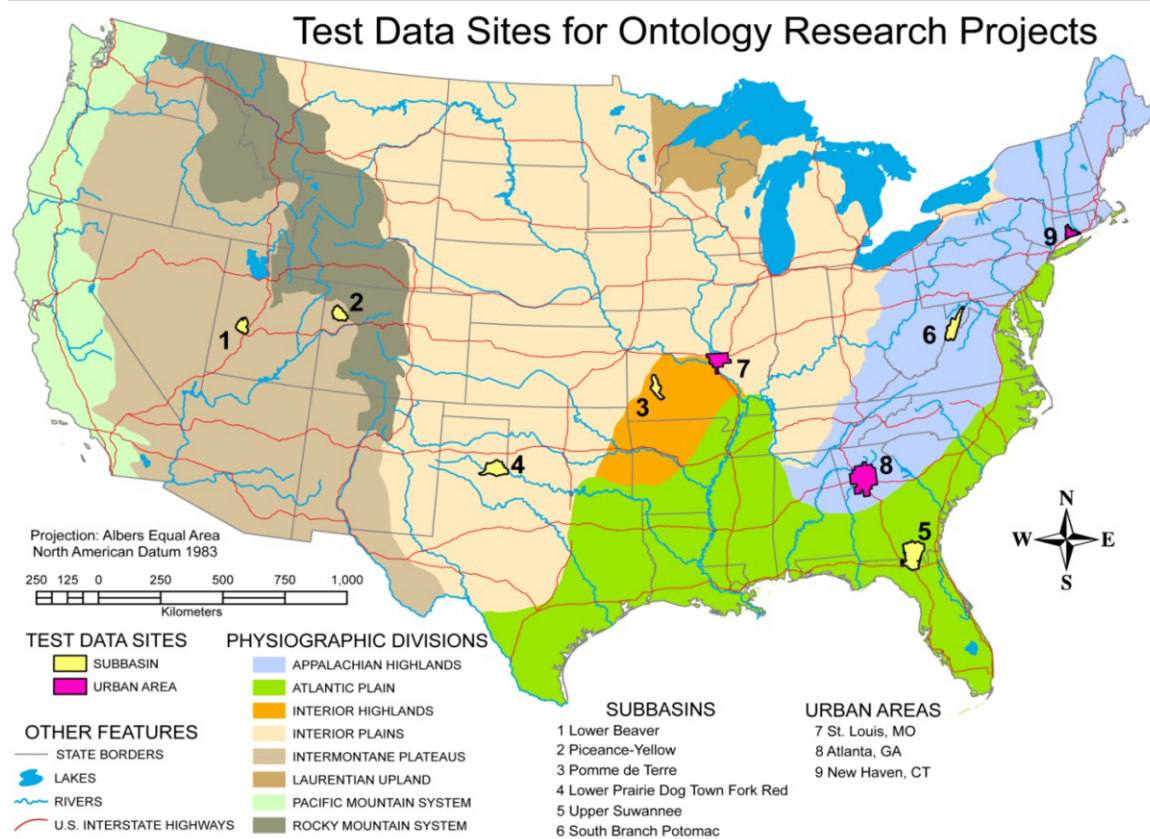

Fig. 2. Location of USGS research datasets for developing ontology and semantics for *The National Map.*

souri; and New Haven, Connecticut, included as an urban coastal site. Each of these test areas includes the eight standard layers of *The National Map*, land cover, structures, boundaries, hydrography, geographic names, transportation, elevation, and orthoimagery (see Fig. 1).

To make USGS data available to the Semantic Web and the Linked Open Data Community, the USGS converted data for the nine research test areas to RDF and GML. Conversion of the sample site datasets to RDF has followed the general approach of defining the subject, predicate, and object of RDF as the feature identifier, feature name or other attribute or relation, and feature instance or object of the relation, respectively. A requirement is the capability to pose SPARQL queries from which results can be graphically displayed on map. Thus, the coordinates must be associated with the RDF resource. This association is done through GML and allows access and use by any traditional program that can process GML. A SPARQL query of the RDF data can retrieve the needed result and the final output can be used to generate a map from the GML coordinate store as needed. All GML entities and operations used in the data conversion and semantic queries follow the OGC standard for GML [28].

In the initial conversion the native format (usually ArcGIS GeoDatabase, [14]) data were converted to GML with each entity possessing a unique identifier. The eight standard topological relations defined by OGC were precomputed from the GML (see Fig. 3 for an example). The feature data were converted from GML to RDF triples maintaining identifiers from the GML.

The required conversion processes and structure of the resulting data with access to the original geometry are different based on the original geometry of the geographic data sources. The following discussion is separated into point, vector, and raster to describe the different processes required for conversion. The structures and geographic names layers use point objects as the geometric base of the data elements. The boundary, hydrography, and transportation layers use vector geometry with point, line, and area objects as the basic data elements. The land cover, elevation, and orthoimage layers use raster geometry with pixels or cells as the basic geometric unit. Objects in the raster layers must be defined and referenced over the cell geometry for access and manipulation.

4.1. Point data

The point datasets for *The National Map* include geographic names and structures. Whereas structures data in *The National Map* will eventually be generated using the polygonal boundary for the structure outline, currently available data use a single point at the proximate center of the building or other structure. Thus, at present structures are converted to RDF using a point geometry model.

The basic conversion for the point data proceeded as follows. Point data for *The National* Map are stored in Esri geodatabase or shape file formats [14]. These files are used to create GML documents to store the geometric data. The output of the conversion process is written to an N3 document [2]. Complete description of this process including conversion from geodatabase, personal geodatabase, and shape files to GML and to N3 is presented in [4].

Each point feature in the Geographic Names Information System (GNIS) is formatted as a name associated with a location. The conversion of this format to RDF triples uses the simple convention that the feature identifier is the object in the RDF triple (Fig. 3). Figure 3 also presents the result in GML including the coordinates for the structure location.

4.2. Vector data

The conversion of vector formatted geospatial data for hydrography, transportation, boundaries, and structures (Table 2) for the test sites to the linked data format of the Semantic Web proceeded with the following general approach. The subject, predicate, object format of RDF for the semantic web was constructed from the entities as defined in formats of *The National Map*. For example, for a stream in the NHD of *The National Map*, flowline is the primary feature of the stream reach that provides connections of the hydrographic network. The subject is

Table 2
Count and volume for converted triples

Dataset	Triple count	File size
Hydrography	20,000.000	2.7 Gb
Transportation	25,000,000	2.4 Gb
Boundaries	52,000	189 Mb
Structures	388,000	37 Mb

Query Text

PREFIX struct: <http://cegis.usgs.gov/rdf/struct#>
PREFIX gt: <http://cegis.usgs.gov/rdf/geometry#>
PREFIX structfid: < http://cegis.usgs.gov/rdf/struct/featureID#>
PREFIX transfid: < http://cegis.usgs.gov/rdf/trans/featureID#>

Select ?name ?gml where
{
structfid:_CT001425 struct:name ?name
structfid:_CT001425 gt:gml ?gml
}

```
<gml:featureMember>
  <ogr:Struct_Point fid="F42">
    <ogr:geometryProperty>
      <gml:Point srsName="EPSG:4326">
        <gml:coordinates>-72.988109997147887,41.242280002650432</gml:coordinates>
      </gml:Point>
    </ogr:geometryProperty>
    <ogr:OBJECTID>43</ogr:OBJECTID>
    <ogr:Permanent_Identifier>19112d9d-f1ad-4f16-b1b9-cde2bd9e57f4</ogr:Permanent_Identifier>
    <ogr:Source_FeatureID>CT001425</ogr:Source_FeatureID>
    <ogr:Source_DatasetID>b934057e-3281-4833-a7ac-ceb9cb12a7c8</ogr:Source_DatasetID>
    <ogr:Source_DataDesc>HAZUS Database / MH Version 1.1, January 2005</ogr:Source_DataDesc>
    <ogr:Source_Originator>FEMA</ogr:Source_Originator>
    <ogr:Data_Security>5</ogr:Data_Security>
    <ogr:Distribution_Policy>E4</ogr:Distribution_Policy>
    <ogr:LoadDate>2005/10/19 0:00:00</ogr:LoadDate>
    <ogr:FType>730</ogr:FType>
    <ogr:FCode>73002</ogr:FCode>
    <ogr:Name>ALMA E. PAGELS SCHOOL</ogr:Name>
    <ogr:IsLandmark>0</ogr:IsLandmark>
    <ogr:PointLocationType>1</ogr:PointLocationType>
    <ogr:AdminType>0</ogr:AdminType>
    <ogr:AddressBuildingName></ogr:AddressBuildingName>
    <ogr:Address>26 BENHAM HILL ROAD</ogr:Address>
    <ogr:City>WEST HAVEN</ogr:City>
    <ogr:State>CT</ogr:State>
    <ogr:Zipcode>06516-6545</ogr:Zipcode>
    <ogr:GNIS_ID></ogr:GNIS_ID>
    <ogr:Foot_ID></ogr:Foot_ID>
    <ogr:Complex_ID></ogr:Complex_ID>
  </ogr:Struct_Point>
</gml:featureMember>
```

Fig. 3. Query text for a structure feature from RDF data of a sample from *The National Map.* The resulting GML from the query is shown in the bottom of the figure.

the feature identifier, in the case of a Flowline, it is the reach code as defined in NHD (fid: 77127453 in Fig. 4). The predicate is the particular property of the flowline being modeled in the triple, its length, for example http://cegis.usgs.gov/rdf/geometry#length.

There are 17 objects and depend on the predicate. For example, the object of the predicate geometry#length is a literal number; the object of geometry#intersects is another flowline. The object of geometry#gml are the coordinates of the flowline. Figure 4 shows a query and the detailed set of flowline characteristics that are the distinct properties or predicates of the flowline. Each subject (reach code identifier) has many distinct predicates and objects

associated with it to capture the stream characteristics. As with the point data, the geometry of the flowline is represented by coordinates stored in GML.

4.3. Raster data

Query and access to raster data on the Semantic Web poses unique problems since geographic features to be represented as ontological objects are not defined in the structure of the data, which is a grid of pixel values or digital numbers. Traditional processing of raster data has treated the entire raster grid as a coverage, as in Web Coverage Services, or

Query Text

```
PREFIXqgis: <http://cegis.usgs.gov/rdf/geometry#>
PREFIXfid:  <http://cegis.usgs.gov/rdf/nhd/featureID#>
PREFIXroad: <http://cegis.usgs.gov/rdf/trans/featureID#>
PREFIXnhd:  <http://cegis.usgs.gov/rdf/nhd#>
PREFIXtrans:<http://cegis.usgs.gov/rdf/trans#>


selectdistinct?predicatewhere{
fid:_77127453?predicate[]


}
```

Result at:

```
http://131.151.2.169:8890/sparql?default-graph-
uri=&query=PREFIX+fid%3A+++%3Chttp%3A%2F%2Fcegis.usgs.gov%2Frdf%2Fnhd%2FfeatureID%23%3E%0D%0A%0D%0Aselect+distinct+%3
Fpredicate+where+{%0D%0Afid%3A_77127453+%3Fpredicate+[]%0D%0A%0D%0A}%0D%0A&format=text%2Fhtml
```

Fig. 4. Query for stream flowline from RDF data. The results are the predicates of the flowline for a sample from the National Hydrography Dataset of *The National Map*.

provided procedures to extract vector objects from the raster matrix. Unless each pixel in a raster data matrix is treated as a separate entity in an ontology, definition of geographic features or ontological objects over the raster grid is required. Although a significant literature exists on image segmentation and object extraction from raster image data (see standard texts on remote sensing and image processing, such as [24]), there has been little work on ontology and semantics with raster geometry. In general, the approach to this problem is first to develop vector objects from image segmentation then use existing methods for building ontology and semantics for the vector objects. However, for relational data, [25] proposed methods to extend the Geographic Structured Query language (GSQL) to support raster data. By defining specific abstract data types (ADTs), such as Pixel, Raster Region, and RasterCoverage and formalizing data objects and operations on these ADTs, GSQL has been extended to query raster objects. [35] also provide an approach to raster data semantics. Their approach is in three stages requiring conceptualization, synthesis, and description of objects in the raster data. Neither of these approaches is directly implemented for the Semantic Web and neither uses an RDF structure for the raster objects.

The raster data layers in *The National Map* are land cover, elevation, and orthographic images (see Fig. 1). Geomorphic entities are typical examples of geographic features dependent on a raster representation. For a specific feature example, this discus-

sion will use the feature crater with the particular feature instance of Meteor Crater (Figs 5a and 5b), Arizona. Note that on a topographic map, Meteor Crater is represented only by a name and the map user must interpret the feature from the extent of the name and contours or from the orthographic image (orthographic images are now included as a layer of US Topo). Thus, a part of the task of representing the crater feature is the definition of its extent in a form a user will understand. Whereas, Meteor Crater is a graphically well-defined feature and easily interpreted by most users from the image or contour map, other geomorphic features, such as hills, are more difficult to identify and have indeterminate boundaries [5].

Unlike other approaches that extract the semantic objects from the raster data, our approach is to determine relevant objects and maintain the raster matrix as the geometric basis of the geographic features of interest. This is essential since a user may want to see a source map or image of the feature in concert with a query result or with other data. This can be understood by examining Meteor Crater as presented in Figs 5a and 5b. A single vector polygon outline of Meteor Crater would not convey the feature characteristics nearly as well as the image or contour map, both of which are raster. The contours could be shown as vector lines and provide the same presentation, but in that case the entities are individual contour lines and not a single entity that is Meteor Crater. The interpretation of the lines as Meteor Crater is again left to the user. Thus, the

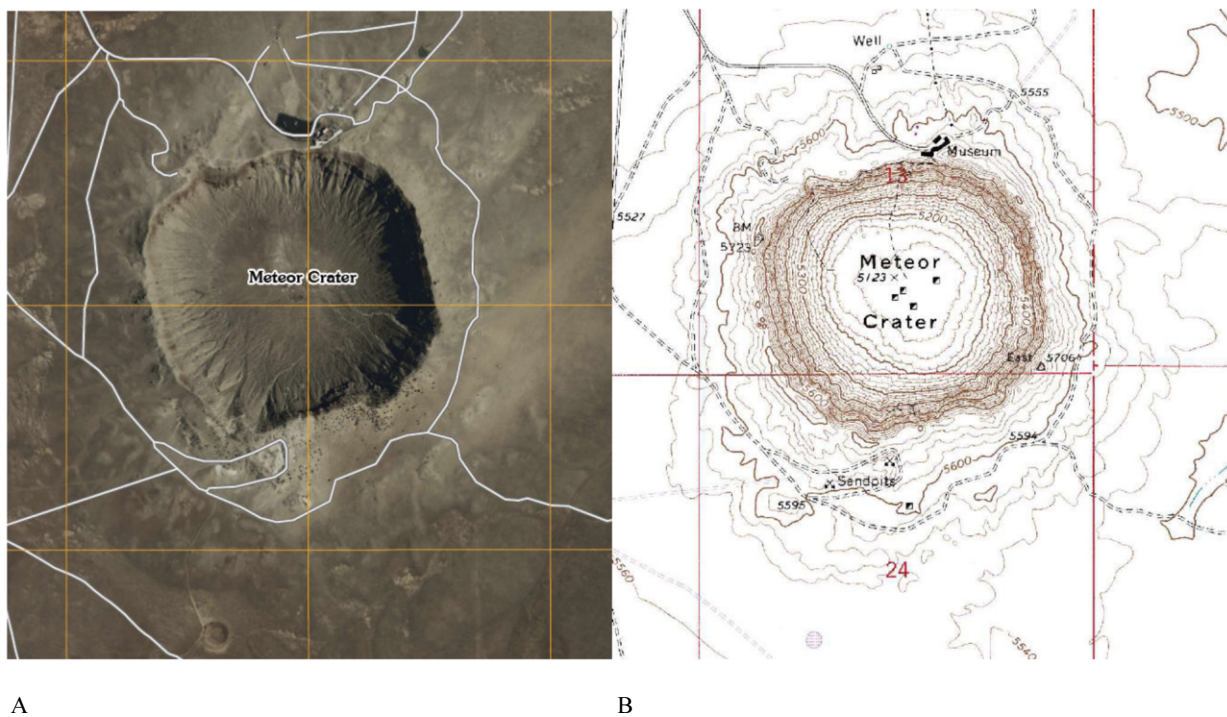A   B

Fig. 5. Orthographic image (A) and topographic map (B) representation of meteor crater.

connection between the ontological object and actual geographical entity in the real world and the raster representation is essential.

The steps involved in the conversion of these types of entities to a semantic representation require that the features be identified in the raster source and a pixel or set of pixels selected as the basic geometric footprint for the feature [48]. This identification results in a single pixel for features that can be treated as point features at the resolution of the raster data. An example is well or spring. A linear set of pixels can be used to represent line types of features, such as roads or rivers, based on size of the feature and resolution of the data. Features that span areas, such as Meteor Crater, require contiguous groups of pixels or in some cases non-contiguous groups of pixels to be identified [46,47]. The identification step must be followed by an identification of the relations of the specified feature to other neighboring features.

The specification of the definition, attributes, and relationships of a feature, a prototype from category theory [23,34,45], provide an ODP, which can be used as a basis for similarity matching to classify and identify features. Such patterns are for actual geographical features and may be used for features represented with vector geometry [51] or raster geometry as in the case of Meteor Crater. For Meteor Crater, the ODP would only include the definitional characteristics appropriate for all craters whereas Table 3 provides the set of attributes and relationships of the particular feature instance. For example, the ODP for the class crater includes the relations: has definition: circular-shaped depression…; has attribute: depth; has attribute: shape, etc. This ODP is generic for all craters since all craters share the definition and all have attributes of depth and shape. Meteor Crater has other attributes and relationships that may not be shared by all craters.

Once the features and relations, as specified in the ontology, are identified, the feature is matched to an existing ODP and additional attributes and relationships are defined for the feature instance, as with Meteor Crater above. The newly defined feature instance is linked to the geometric pixel patterns of the raster image. At this point an RDF structure can be created for the feature. Similar to the representation of point and vector data in RDF above, the conversion of the feature and relations to RDF is performed and the raster geometry, pixel, linear set of pixels, or pixel aggregation, is structured in GML, using the GML coverage. To define

Table 3

Meteor crater attributes and relationships

Feature	Crater				
Definition	Circular-shaped depression at the summit of a volcanic cone or				
	one on the surface of the land caused by the impact of a meteorite;				
	a manmade depression caused by an explosion (caldera, lua).				
Instance	Meteor Crater				
	GNIS ID 7945				
Attributes					
	Location	UTM	E 497,959.94 m	N 3,876,020.68 m	Zone 12
		PLSS	T 19 N, R 12 1/2 E, Section 13 and 24		
		MBR	Max E 498,536.79 m		Min E 497,317.62 m
			Max N 3,876,632.29 m		Min N 3,875,479.58
	Elevation		High	5,723 ft	
			Low	5,123 ft	
	Depth		600 ft		
	Shape	Circular			
			Inner Diameter	0.50 mi (0.833 km)	
			Outer Diameter	0.75 mi (1.25 km)	
	Rim width		0.125 mi (0.2 km)		
	Contour at outer perimeter		5,600 ft		
	Contour at inner perimeter		5,180 ft		
Relationships					
	Surrounded by roads				
	Adjacent to Museum		Museum Name:	Meteor Crater Museum	
	Near sand pits				
	Near well				
	Benchmarks on crater		BM 5723	BM East 5706	

the gml:Grid element, a minimum bounding rectangle (MBR) is used for the feature since at this point GML does not allow storage of pixels in other than a rectangular fashion. Eventually, the exact set of pixels that represent a line or polygon will *be* stored, but currently to remain within the GML standard, only the MBR is used.

5. Access to USGS RDF data for research test sites of *The National Map*

To provide access to the research test data converted to RDF, the USGS established a server accessible to the public (http://usgs-ybother.srv.mst.edu). On this server users external to the USGS Intranet can access and download the data in the original Esri and image formats (Geodatabase, shapefile, TIFF) of *The National Map* or in RDF. The USGS has also established a SPARQL Endpoint at http://usgs-ybother.srv.mst.edu:8890/sparql that allows direct query of the data using SPARQL. To illustrate the use of the SPARQL Endpoint, the USGS implemented the relations standardized by OGC from the 9-intersection model [13]. An example relation illustrating a use case with the SPARQL Endpoint and the converted data is shown below. The relation is touches and the use case is "For a given feature, find all other features that touch the given feature." (Fig. 6). Placing the query in the geographic space of data from *The National Map*, it can be phrased about a specific feature: "Find all the tributaries of West Hunter Creek". The result is a series of URIs and when the coordinates from GML of the result are placed on a background map, the graphic in Fig. 7 is the result.

The current capabilities of the endpoint are restricted to the precomputed relationships provided and the values included from the native datasets. For example, one can ask "Which features intersect any feature with the NHD reach code X?" and receive a correct result. However, one could not ask "Which

OpenLink Virtuoso SPARQL Query

Welcome to the SPARQL endpoint for triple data derived from *The National Map* of the U.S. Geological Survey (USGS). The National Map data in USGS distribution formats are available at http://nationalmap.gov/

A graphical, web-based query editor is available here.

The geographical extents represented by the datasets include six watershed areas, two urban areas, and one coastal area.

- Pomme de Terre, MO
- Upper Suwannee, GA-FL
- Lower Prairie Dog Town Fork of the Red, TX
- Lower Beaver, UT
- South Branch of the Potomac, WV
- Piceance-Yellow, CO
- Atlanta, GA
- St. Louis, MO
- New Haven, CT

Please send comments to us at: dvaranka@usgs.gov

The following are possible graph URIs, followed by brief descriptions of their content.

- http://cegis.usgs.gov/rdf/Dec/10/2010/
  - This graph contains data for all nine areas above.
  - The features and their geometries are combined into one resource
  - This graph is no longer being modified so any results from queries should never change.

- http://cegis.usgs.gov/rdf/geosparqltest/
  - This graph contains data for Atlanta, GA only.
  - The features and their geometries are seperated into individual resources as per the GeoSPARQL draft standard
  - This graph may be modified by either fixing any detected errors in the data or adding the remaining areas.

- http://cegis.usgs.gov/rdf/ontologytest/
  - This graph contains data for Pomme de Terre, MO only.
  - The features and their geometries are seperated into individual resources as per the GeoSPARQL draft standard
  - This graph is actively being modified, including adding, removing, and modifying the data and its structure.

Query

Default Graph URI
http://cegis.usgs.gov/rdf/ontologytest/

PREFIX ogc: <http://www.opengis.net/rdf#>
PREFIX fid: <http://cegis.usgs.gov/rdf/nhd/featureID#>

SELECT ?feature ?type
WHERE {
fid:_102217454 ogc:hasGeometry ?geo1.
?geo1 ogc:touches ?geo2.
?feature ogc:hasGeometry ?geo2.
?feature a ?type }

Fig. 6. Initial screen accessed on the USGS SPARQL endpoint with example query using relation touches.

features inside rectangle R have reach code X?" because the rectangle R isn't a precomputed relationship and isn't stored as a predicate. We continue to refine our conversion processes and expand the capabilities of the RDF data. Our current research is to eliminate the precomputation in the conversion and rely on the ontology with defined relationships to drive the query processing. We anticipate applications of these data in environmental modeling and graphical display of model results.

6. Conclusions

The USGS is researching the capabilities of the Semantic Web for supporting query and analysis of geographic data from *The National Map*. As a part of that research, point and vector data for nine re-search test areas have been converted to RDF and made available to the public. A vocabulary of topographic terms has been developed to form the basis for ontology for *The National Map*. To support user interaction with the converted data, the USG provides access for download of the research test data in original formats of the *The National Map*, RDF formatted data, and a SPARQL Endpoint for direct query of the data. The USGS is participating with these data in testing the evolving GeoSPARQL standard and providing methods for users to semantically interact with the data.

Raster data representation on the Semantic Web requires constructing object representations and developing the complete set of attributes and relationships that comprise the ontology for the entities while maintaining the pixel geometry for user access. Approaches to date have relied on conver-
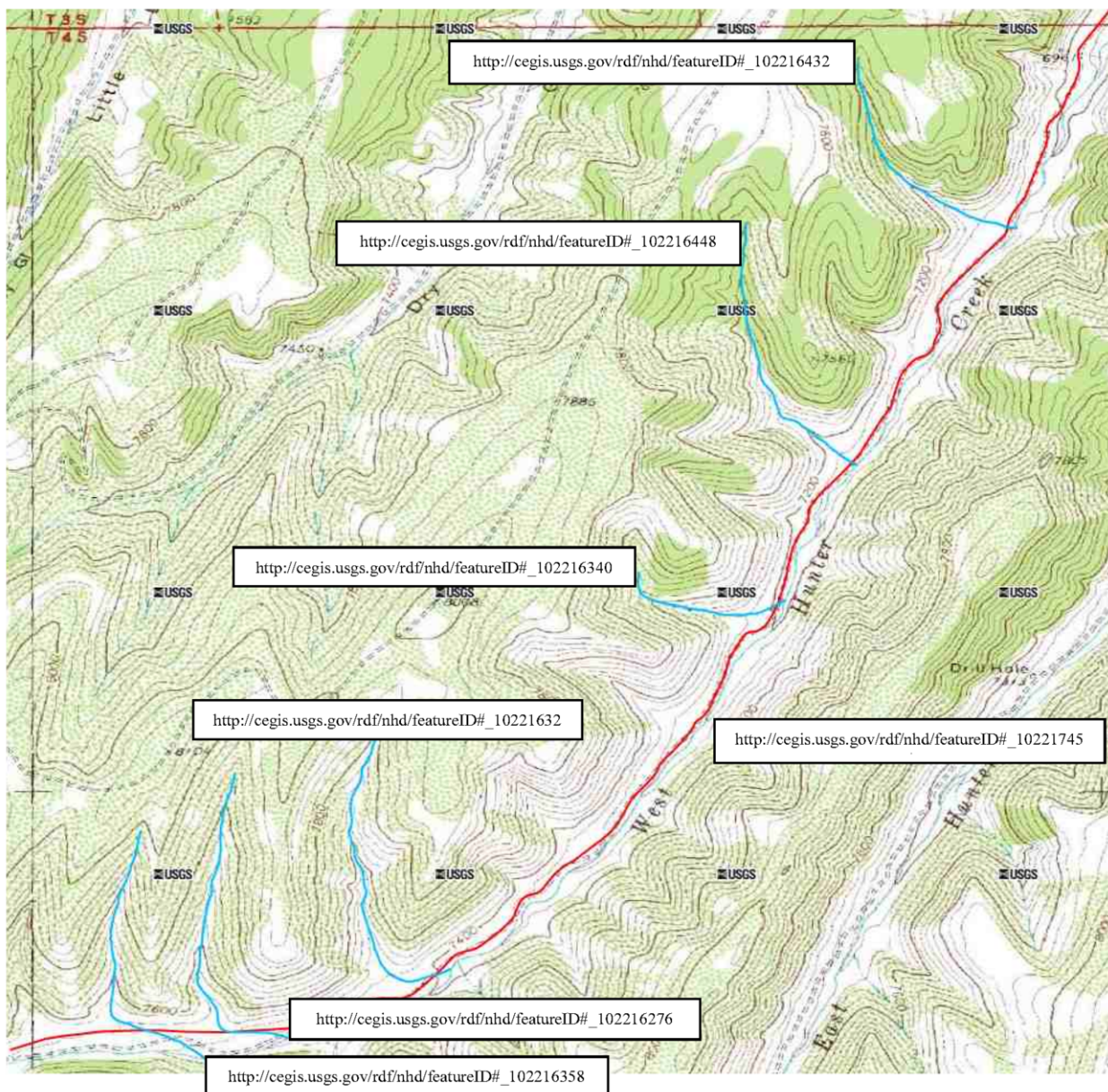
Fig. 7. Graphical result of the query in Fig. 6. West Hunter Creek (http://cegis.usgs.gov/rdf/nhd/featureID#102217454) is shown in red and its tributaries are shown in blue with associated URIs. The background image is a standard USGS Digital Raster Graphic for the quadrangle that includes West Hunter Creek, Colorado.

sion from raster to vector geometry thus losing the original geometric source of the data. The USGS approach is to maintain the pixel structure of the entity from the raster image and build ontology from ODP and specific feature instance attributes and relationships.

References

[1] S. Auer, J. Lehmann and C. Stadler, LinkedGeoData. Agile Knowledge Engineering and Semantic Web (AKSW), University of Leipzig, 2011, accessed August 2, 2011, at URL http://linkedgeodata.org.

[2] T. Berners-Lee, Primer: Getting into RDF & Semantic Web using N3: v. 1.61, 2005, World Wide Web Consortium, accessed March 16, 2011, at http://www.w3.org/2000/10/swap/Primer.

[3] B. Brodaric, A foundational framework for structuring geographical categories, in: *Proc. of First International Workshop on Informational Semantics and Its Implications for Geographical Analysis, Geoscience 2008*, Park City, Utah, September 23, 2008, http://cogsci.uni-osnabrueck.de/~isga08/Brodaric.pdf, 2008.

[4] A. Bulen, J. Carter and D. Varanka, A Program for the Conversion of *The National Map* Data From Proprietary Format to Resource Description Framework (RDF), U.S. Geological Survey Open-File Report 2011-1142, 9 p., http://pubs.usgs.gov/of/2011/1142/, 2011.

[5] P.A. Burrough and A.U. Frank, eds, *Geographic Objects with Indeterminate Boundaries*, Taylor and Francis, London, 1996, p. 352.

[6] G. Câmara, A.M. Vieira-Monteiro, J. Paiva and R.C.M. de-Souza, Action-driven ontologies of the geographical space: Beyond the field-object debate, in: *Proc. of GIScience 2000-First International Conference on Geographic Information Science*, Savannah, GA, M.J. Egenhofer and D.M. Mark, eds, 2000, pp. 52–54.

[7] H.K. Caro and D.E. Varanka,. Analysis of Spatial Relation Predicates in U.S. Geological Survey Feature Definitions. U.S. Geological Survey Open-File Report 2011-1235, 2011, 37 p., http://pubs.usgs.gov/of/2011/1235.

[8] R. Casati and A. Varzi, *Parts and Places, the Structures of Spatial Representation*, The MIT Press, Cambridge, MA, 1999.

[9] H. Couclelis, Ontologies of geographic information, *International Journal of Geographical Information Science* **24**(12) (2010), 1785–1809.

[10] D.J. Dean, Characterizing spatial databases via their derivation: A complement to content ontologies, *Transactions in GIS* **11**(3) (2007), 399–412.

[11] C. Dolbear, G. Hart and J. Goodwin, What OWL has done for geography and why we don't need it to map read, in: *Proc. of the OWL Experiences and Directions (OWLED) 2006, Workshop*, Athens, Georgia, USA, B. Cuenca, P. Hitzler, C. Shankey and E. Wallace, eds, CEUR Workshop Proceedings, Vol. 216, 2006, CEUR-WS.org.

[12] C. Dolbear and J. Goodwin, Position paper on expressing relational data as RDF, in: *Proc. of W3C Workshop on RDF Access to Relational Databases, 25–26 October, 2007, W3C*, Cambridge, MA, USA, http://www.w3.org/2007/03/RdfRDB/papers/dolbear.pdf, 2007.

[13] M.F. Egenhofer and R.D. Franzosa, Point set topological spatial relations, *International Journal of Geographical Information Systems* **5**(2) (1991), 161–174.

[14] Esri, Geodatabases, Esri, accessed March 30, 2011, at http://resources.arcgis.com/content/geodatabases/10.0/about, 2011.

[15] A. Gangemi, Ontology design patterns for Semantic Web content, in: *Proc. of the Fourth International Semantic Web Conference*, M. Musen, et al., eds, Lecture Notes in Computer Science, Vol. 3729/2005, Springer, Berlin, 2005, pp. 262–276.

[16] A. Gangemi and V. Presutti, Ontology design patterns, in: *Handbook of Ontologies*, 2nd edn, S. Staab and R. Studer, eds, Springer, Berlin, 2009.

[17] N. Guarino and C. Welty, Evaluating ontological decisions with OntoClean, *Communications of the ACM* **45**(2) (2002), 61–65, ACM Press, New York.

[18] S. Hahmann and D. Burghard, Connecting LinkedGeoData and geonames in the spatial semantic web, in: *Proc. of GIScience 2010 Extended Abstracts,* Zurich, Switzerland, R. Purves and R. Weibel, eds, 2010, pp. 28–34.

[19] G. Hart, J. Goodwin and T. Pehle, GeoVoCampsSouthampton201, 2011, accessed Nov. 10, 2011, at: http://vocamp.org/wiki/GeoVoCampSouthampton2011.

[20] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, 1st edn, J. Hendler and F. van Harmelen, eds, Synthesis Lectures on the Semantic Web: Theory and Technology, Vol. 1, No. 1, Morgan & Claypool, 2011, pp. 1–136, accessed May 31, 2011, at http://linkeddatabook.com/editions/1.0/.

[21] M. Kokla and M. Kavouras, Fusion of top-level and geographical domain ontologies based on context formation and complementarity, *International Journal of Geogrpahical Information Science* **15**(7) (2001), 679–687.

[22] W. Kuhn, Geospatial semantics: Why, of what, and how?, in: S. Spaccapietra and E. Zimányi, eds, *Journal on Data Semantics III, Lecture Notes in Computer Science* **3534**(3) (2005), 1–24.

[23] G. Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*, University of Chicago Press, Chicago, 1987, p. 602.

[24] T.M. Lillesand, R.W. Kiefer and J. Chipman, *Remote Sensing and Image Interpretation*, 6th edn, John Wiley and Sons, New York, 2007, 804 p.

[25] Y. Liu, Y. Lin, S. Qin, Y. Zhang and L. Wu, Research on GSQL estension supporting raster data, *Journal of Image and Graphics* (2005), accessed May 31, 2011, at http://en.cnki.com.cn/Article_en/CJFDTOTAL-ZGTB20050100J.htm.

[26] C. Masolo, S. Borgo, A. Gangemi, N. Guarino and A. Oltramari, WonderWeb Deliverable D18, Ontology Library (final). Laboratory for Applied Ontology, 2003, http://www.loa-cnr.it/Papers/D18.pdf.

[27] OGC, Open Geospatial Consortium, Inc., 2010, accessed March 25, 2010, at: http://www.opengeospatial.org/.

[28] OGC, *Geography Markup Language (GML) Encoding Standard*, Open Geospatial Consortium, Inc., 2011, accessed May 31, 2011, at http://www.opengeospatial.org/standards/gml.

[29] Ontology Engineering Group (OEG). GeoLinked Data. Universidad Politecnica de Madrid, accessed August 2, 2011, at http://geo.linkeddata.es, 2011.

[30] Ordnance Survey (OS), Ordnance Survey Ontologies. Accessed August 8, 2011, at: http://www.ordnancesurvey.co.uk/oswebsite/ontology/, 2011.

[31] T. Pehle, NeoGeoVoCamp Summary Report. Neogeo, 2009, accessed August 9, 2011, at: http://sites.google.com/site/neogswvocs/.

[32] M. Perry and J. Herring, eds, *GeoSPARQL – A Geographic Query Language for RDF Data*, Open Geospatial Consortium Inc., 2011, project document reference number OGC 09-157-r1.

[33] C. Portele, *OpenGIS Geography Markup Language (GML) Encoding Standard, v. 3.2.1*, Open Geospatial Consortium, Inc., 2007, OGC 07–036, accessed March 30, 2011, at http://www.opengeospatial.org/standards/gml.

[34] E. Rosch, Principles of categorization, in: *Cognition and Categorization*, E. Rosch and B.B. Lloyd, eds, Halstead Press, New York, 1978, pp. 27–48.

[35] R. Quintero, M. Torres, M. Moreno and G. Guzman, Towards a semantic representation of raster spatial data, in: *Proc. of GeoSpatial Semantics, Third International Conference, GeoS 2009*, Mexico City, Mexico, December 2009, LNCS, Vol. 5892, Springer, 2009, pp. 63–82.

[36] J.M. Salas and A. Harth, NeoGeoVocabulary: Defining a shared RDF representation for GeoData, 2011, accessed August 9, 2011, at: http://geovocab.org/doc/survey.html.

[37] Spatial Data Transfer Standard Technical Review Board, 1997, Spatial Data Transfer Standard (SDTS) – Part 2, Spatial Features; Draft for Review, Federal Geographic Data Committee.

[38] S. Sen, Two types of hierarchies in geospatial ontologies, in: *Proc. of GeoS 2007*, F. Fonseca, M.A. Rodrigues and S. Levashkin, eds, LNCS, Vol. 4853, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 1–19.

[39] M.K. Smith, C. Welty and D.L. McGuinness, OWL Web Ontology Language. W3C Recommendation 10 February 2004. http://www.w3.org/TR/2004/REC-owl-guide-20040210/#FunctionalProperty, 2009.

[40] L. Sugarbaker, K. Coray and B. Poore, *The National Map* Customer Requirements: Findings from Interviews and Surveys. USGS Open-File Report 2009–1222, USGS, Reston, VA, 2009.

[41] U.S. Board on Geographic Names, Geographic Names Information System, 2010, accessed March 25, 2010, at: http://geonames.usgs.gov/domestic/index.html.

[42] U.S. Geological Survey, National Geospatial Program Standards, 2001, accessed March 23, 2010, at http://rockyweb.cr.usgs.gov/nmpstds/dlgstds.html.

[43] U.S. Geological Survey, National Hydrography Dataset (NHD) Model (v 2.0), 2010, accessed August 15, 2011, at: http://nhd.usgs.gov/NHDv2.0_poster_6_2_2010.pdf.

[44] U.S. Geological Survey, A Topographic Feature Vocabulary for Geospatial Ontology Development. U.S. Geological Survey, 2011, http://cegis.usgs.gov/ontology.html.

[45] E.L. Usery, Category theory and the structure of features in geographic information systems, *Cartography and Geographic Information Systems* **20**(1) (1993), 5–12.

[46] E.L. Usery, Implementation constructs for raster features, in: *Proc. American Society for Photogrammetry and Remote Sensing Annual Convention*, *ASPRS*, Reno, Nevada, Bethesda, MD, 1994, pp. 661–670.

[47] E.L. Usery, Display of geographic features from multiple image and map databases, in: *Proc. of International Society for Photogrammetry and Remote Sensing, Commission IV Symposium on Mapping and Geographic Information Systems, International Archives of Photogrammetry, Volume XXX, Part B4*, Athens, G, 1994, pp. 1–9.

[48] E.L. Usery, A feature-based geographic information system model, *Photogrammetric Engineering and Remote Sensing* **62**(7) (1996), 833–838.

[49] D. Varanka, Landscape features, technology codes, and semantics in U.S. national topographic mapping databases, in: *Proc. of the International Conference on Advanced Geographic Information Systems & Web Services (GEOWS)*, Cancun, Mexico, February 1–7, 2009, 2009.

[50] D. Varanka, A topographic feature taxonomy for a U.S. national topographic mapping ontology, in: *Proc. of International Cartographic Conference*, Santiago, Chile, ICA CD-ROM publication, 2009.

[51] D. Varanka, Ontology patterns for complex topographic feature types, *Cartography and Geographic Information Science* **38**(2) (2011), 126–136.

[52] D. Varanka and E.L. Usery, Special section: Ontological issues for *The National Map*, *Cartographica: The International Journal for Geographic Information and Visualization* **45**(2) (2010), 103–104.

[53] C. Welty and N.Guarino, Supporting ontological analysis of taxonomic relations, *Data and Knowledge Engineering* **39** (2001), 51–74.

[54] W3C 2004. Resource Description Framework (RDF), accessed May 31, 2011, at http://www.w3.org/RDF/.

[55] W3C 2011. Semantic Web, accessed May 31, 2011, at http://www.w3.org/standards/semanticweb/.