



GNIS-LD: Serving and Visualizing the Geographic Names Information System Gazetteer as Linked Data

Blake Regalia^{1(✉)}, Krzysztof Janowicz^{1(✉)}, Gengchen Mai¹, Dalia Varanka²,
and E. Lynn Usery²

¹ STKO Lab, University of California, Santa Barbara, USA
{blake, gengchen.mai}@geog.ucsb.edu, janowicz@ucsb.edu

² U.S. Geological Survey, Reston, USA
{dvaranka, usery}@usgs.gov

Abstract. In this dataset description paper we introduce the GNIS-LD, an authoritative and public domain Linked Dataset derived from the Geographic Names Information System (GNIS) which was developed by the U.S. Geological Survey (USGS) and the U.S. Board on Geographic Names. GNIS provides data about current, as well as historical, physical, and cultural geographic features in the United States. We describe the dataset, introduce an ontology for geographic feature types, and demonstrate the utility of recent linked geographic data contributions made in conjunction with the development of this resource. Co-reference resolution links to GeoNames.org and DBpedia are provided in the form of owl:sameAs relations. Finally, we point out how the adapted workflow is foundational for complex Digital Line Graph (DLG) data from the USGS National Map and how the GNIS-LD data can be integrated with DLG and other data sources such as sensor observations.

Keywords: Linked data gazetteer
Geographic Names Information System · Geospatial semantics
Feature type ontology

Resource type: Dataset

Permanent URL: <http://w3-id.org/usgs>.

1 Introduction and Motivation

Linked Data and Semantic Web technologies are very popular in the broader Geosciences as they address several key challenges [12] within those domains such as improving interoperability across heterogeneous datasets, e.g., spanning physical and human geography, easing the publishing and retrieval of datasets, support co-reference resolution without enforcing global consistency, and so forth. However, similar to many technologies before, the early Linked Data cloud faced

a chicken-and-egg problem. The value proposition of Linked Data and Semantic Web technologies became evident to industry, government agencies, and end users, only after a substantial number of datasets were deployed, interlinked, and made accessible using query endpoints, graphical user interfaces, and services, such as question answering. To overcome this challenge, the early Linked Data cloud was driven by Semantic Web researchers triplifying popular, third-party datasets. While this rapidly growing size of data sources helped fuel the initial enthusiasm for Linked Data and showcase interesting applications, it was not without its own shortcomings.

For instance, datasets were triplified, and ontologies were created without substantial domain expertise, and the published datasets and their endpoints were not maintained [9]. At one point, according to <http://sparqls.ai.wu.ac.at/>, 54% of monitored endpoints had an uptime of 0–5%. This is not surprising as university projects are often not well suited for long-term maintenance, quality control, end-user support, and other tasks that do not align with the research and innovation focus of universities. The original data providers, such as government agencies, research centers, and the industry, however, did not yet have the interest and expertise to deploy their data as Linked Data. Nonetheless, these early datasets (and vocabularies) served their purpose, namely showcasing the potential of Linked Data and overcoming the chicken-and-egg problem.

Thanks to these initial datasets, we are currently witnessing a second wave of Linked Data publishing, namely one driven by the providers themselves such as research libraries, government agencies, large-scale data infrastructures, e.g., in the context of NSF’s EarthCube effort, and industry. These efforts often require specific strategies, workflows, and tools to ensure long-term maintenance and support for their specific target audience. In contrast to individual research projects, these larger endeavors are only launched when the responsible organizations are convinced that they can be kept alive on the long term. Among many other factors, this requires technology transfer between research, industry, and government agencies [6], customization of (open source) software to internal workflows, strategies for long-term maintenance and (continuous) release cycles, as well as administration and support. The resulting linked datasets are not meant to replace existing linked datasets but to complement them by providing an authoritative alternative.

Example of domains in which this second wave of Linked Data publishing is currently ongoing are the Earth Sciences and Geography. To give a concrete example, the GeoNames gazetteer is one of the most interlinked datahubs on the Linked Data cloud. GeoNames ingests several data sources and mixes authoritative data (e.g., from Geographic Names Information System) with volunteered geographic information (VGI). However, it does not maintain a SPARQL endpoint, does not make use of `rdf:type` predicates but uses its own `gn:featureCode` property instead, introduces its own feature type catalog that is not used by any other geographic data set, and only contains a subset of the data made available by GNIS. It does, however, introduce a vast variety of geographic features from other (volunteered) resources. Consequently, it is desirable

to complement GeoNames with authoritative data sources that are produced and maintained by the organizations responsible for the data. This way, different target audiences can prioritize their needs, e.g., in terms of endpoint availability, update intervals, coverage, accuracy, and so forth.

The GNIS gazetteer is an essential, authoritative dataset across domains and tasks as places in general act as nexuses that connect actors, events, and objects. To give but a few examples, exhibits such as photographs and paintings can depict a location and are taken at a location. Specimen and samples more generally are collected at a specific location and stored at another one. Agencies and news organizations need to make sure that they refer to the same location despite multiple places sharing the same name or using different spelling variants.

In this paper, we will introduce GNIS-LD, an authoritative Linked Data version of the Geographic Names Information System¹. We will discuss its value as a testbed for future linked geographic data aimed at supporting the scale and geometric complexity of very large geographic information repositories. We will discuss the need for complementing GeoSPARQL with dereferenceable URIs and geometric metadata [19] and for serving the dataset in a client-sided, extensible Semantic Web Browser [18]. As for describing the dataset itself, we will also introduce an ontology for geographic feature types based on the Enhanced Digital Line Graph Design specs [8] used in the GNIS and the USGS National Map as well as a co-reference resolution graph between GNIS, DBpedia, and GeoNames. Finally, we will show an example for integrating GNIS-LD with Digital Line Graph data about waterbody segments with sensor stations that measure properties such as flow velocity. Our work follows the tradition of other geographic data source providers such as [1].

2 Geometry and the Linked Data Web

Answering the need to store and query geospatial data on the Semantic Web, OGC's GeoSPARQL [16] addressed the most pertinent issues surrounding alternative approaches at the time. While the proposed standard has been foundational in establishing Linked Data as a compatible publishing mechanism for traditional Geographic Information Systems (GIS), it has also revealed major limitations in practice [2]. Most notably, the need to serialize complex geometries as RDF literals has bogged down the storage, transmission, and query potential that traditional GIS have been refining for decades.

More recently, there has been interest in mitigating the considerable storage and query impact that accompanies implementations of the GeoSPARQL standard. For instance, Debruyne et al. [7] curbs geospatial processing demands by storing several copies of a feature's geometry at different levels of polygon simplification. Bereta et al. [3] avoid the need to store geometry data in a hefty serialization format that normally persists in a triplestore's RDF literal bank by instead bridging relational spatial databases with SPARQL engines, allowing

¹ Throughout the paper, we will use prefixed names for IRIs and provide the mappings here: <http://phuzzy.link/context/gnis-ld>.

geometry to persist in their native GIS (which internally is stored in some binary geospatial format) while virtualizing the existence of a GeoSPARQL-compatible serialization format such as Well Known Text (WKT) to the end user.

Our approach is to complement GeoSPARQL's strengths and overcome its limitations by rethinking the need for storing or virtualizing geometry data in the triplestore entirely, especially considering that GeoSPARQL implementations *already* depend on auxiliary binary geometry objects for geospatial query processing. As previously described [19], it is important to recognize that the main explanation for retaining a human-readable serialization of complex geometries in a triplestore (over the alternative) is so that SPARQL query results may transmit geometry data. However, complex geometries are not human-readable anyway as they consist of hundreds or thousands of coordinate pairs. Therefore, we suggest that geographic linked data publishers use dereferenceable URIs to represent complex geometric objects instead. Using a named node in this capacity means that each geometric object has its own URI as opposed to the common blank-node approach often used in the wild with GeoSPARQL objects. It's important to note that we also encourage adding triples to each geometric object to describe it, such as the feature's centroid, its bounding box, digitizing scale, and so forth. The contents of the geometry are then accessible by dereferencing the URI, allowing the data to persist in a native GIS on the host, or even remotely on another source which greatly improves the reusability of geometry data on the Linked Data Web as a whole.

This approach has been instrumental in meeting the storage, transmission, and query demands seen at the scale of the USGS datasets from the National Map, which includes a comprehensive coverage of the topography and water features throughout the entire United States. These datasets contain hundreds of thousands of complex geometries such as high-resolution polylines and polygons. In Listing 1, we show an excerpt from the extended dataset for two features that have a geometry. The first feature's geometry is a point which is accompanied by its complete WKT literal, while the second feature's geometry is a *linestring* with a WKT literal for its bounding box. Both geometry URIs can be dereferenced to obtain their full, encapsulated geometry data in a serialization format determined by the client via content negotiation. Together with the dereferencing functionality provided by the server, GNIS-LD passes all tests on Vafu (and other Linked Data validators)^{2, 3}.

The client may use content negotiation on a dereferenceable URI to download a feature's geometry data in a serialization format that suits their needs. For our particular implementation, these HTTP requests are handled by the server⁴ that queries a local geodatabase in order to extract and convert a feature's geometry into the format given by the request's 'Accept' header. A few example requests are shown in Listing 2.

² See, for instance, <https://bit.ly/2G7anvR>.

³ <http://phuzzy.link/browse/gnis-ld.org/sparql/select#usgeo-multipolygon:nhd.44560536>.

⁴ <https://github.com/blake-regalia/gnis-ld.org/blob/master/lib/server/server.js>.

```

# GNIS features have point geometries. Geometry URIs dereference to their coordinate data
gnisf:430 ago:geometry usgeo-point:gnisf.430 .

# The named node for a geometry also has triples, such as its WKT serialization
usgeo-point:gnisf.430 geosparql:asWKT
  ↳ "<http://www.opengis.net/def/crs/OGC/1.3/CRS84>POINT(-113.667433
  ↳ 34.38335)"^^geosparql:wktLiteral .

# NHD features have complex geometries. They can also be dereferenced
nhd:44560536 ago:geometry usgeo-multipolygon:nhdf.44560536 .

# A geometry includes triples such as how many points it has, its bounding box, and so forth
usgeo-multipolygon:nhdf.44560536 rdf:type ago:MultiPolygon ;
ago:pointCount 9059 ; ago:ringCount 225 ;
ago:centroid "<http://www.opengis.net/def/crs/EPSSG/0/4326>POINT(-120.033432894877
  ↳ 39.0959954262897)"^^geosparql:wktLiteral ;
ago:boundingBox "<http://www.opengis.net/def/crs/EPSSG/0/4326>POLYGON((-120.164012990588
  ↳ 38.9283332062421,-120.164012990588 39.2489710724111,-119.926092324291
  ↳ 39.2489710724111,-119.926092324291 38.9283332062421,-120.164012990588
  ↳ 38.9283332062421))"^^geosparql:wktLiteral ; # and more ...

```

Listing 1 Two geographic features (one located by a point, the other by a linestring) and the triples describing their geometric representations.

```

#!/bin/bash
feature="http://gnis-ld.org/lod/geometry/multipolygon/nhdf.44560536"

# request geometry as GeoJSON
curl "$feature" -H "Accept: application/vnd.geo+json"
## {"type":"MultiPolygon","coordinates":[[[[-119.989951524192,...]]]]}

# request geometry as Well-Known Text
curl "$feature" -H "Accept: text/plain"
## MULTIPOLYGON((-119.989951524192 39.2487662057447,...))

```

Listing 2 Example curl requests demonstrating content negotiation for various serialization formats when downloading a geometry's data.

3 Converting GNIS to Linked Data

USGS/BGN maintains the official GNIS in several relational database tables which get published regularly in data dumps as flat CSV files⁵. The contents of the GNIS include national features and topical gazetteers, which primarily contain records that represent the naming of physical or cultural places on the surface of the Earth. Each entry has various attributes such as the type of geographic feature it represents, its WGS84 point coordinate, the city, county and state it belongs to, the elevation above sea level, the date the entry was created, the original map source, alternative names, historical records, and an official citation.

Our process begins at these data dumps, which we feed through a collection of scripts⁶ that transform the CSV files into RDF by following steps derived from the GNIS topical gazetteer schema⁷. We introduce a simple vocabulary⁸ to

⁵ https://geonames.usgs.gov/domestic/download_data.htm.

⁶ <https://github.com/blake-regalia/usgs-triplifier/tree/master/lib/gnis>.

⁷ https://geonames.usgs.gov/domestic/gaz_fileformat.htm.

⁸ <http://gnis-ld.org/lod/gnis/ontology/>.

Table 1. Dataset statistics

Quantity	Attribute	Instance Count	Feature Type
37,170,932	triples	231,967	Church
6,743,075	named subject nodes	231,508	Stream
2,276,454	GNIS features	216,473	School
2,014,638	sameAs relations to GeoNames	201,066	Populated Place
503,171	sameAs relations to DBpedia	162,509	Locale
494,757	alternative names	160,298	Building
66	feature types	145,534	Cemetery
2,537 MiB	worth of Turtle files	74,282	Reservoir
		70,619	Summit

(a) General statistics about the GNIS-LD dataset.**(b)** Top 10 GNIS feature types by instance counts.

describe GNIS feature attributes and a revised USGS ontology⁹ to describe the feature type class hierarchy and to support the linking of features across datasets, such as those datasets found in The National Map¹⁰. Furthermore, metrics such as elevation above sea level, and length or area of geometric objects, are encoded as XSD-datatype QUDT¹¹ objects.

URIs are minted according to the ID fields that act as foreign keys in relational joins, e.g., a reference to a GNIS feature with ID 1654975 becomes ‘`gnisf:1654975`’. These URIs reflect the *permanent identifiers* assigned by the USGS and so they are guaranteed to always reference the same feature in all versions, i.e., past, present and future, of the GNIS. We also provide `owl:sameAs` links¹² to GeoNames.org, which includes the GNIS as one of its sources (more on that in Sect. 4). However, GeoNames.org does not track the provenance of its features, such as by storing the source id along with a feature’s attributes, so we resort to aligning the GNIS-LD with GeoNames.org by matching exact names, comparing their alternative names, and testing that their locations exist within some distance threshold¹³. This approach may miss matches that have undergone name changes between the two versions of the GNIS. To this end, future work will employ spatial signatures [21] to improve the alignment with GeoNames. Out of the 2.23 million US features on GeoNames.org, we are able to match 90.1% of these records to the GNIS-LD. Alignment with DBpedia also uses exact name string matching but it additionally compares attributes such as the county, state, and place type for each feature. We then use the results from the GeoNames.org matching process to enhance our alignment with DBpedia via `owl:sameAs` transitivity. The number of matches can be seen in Table 1.

⁹ <http://gnis-ld.org/lod/usgs/ontology/>.

¹⁰ <https://viewer.nationalmap.gov/basic/>.

¹¹ <http://www.qudt.org/release2/qudt-catalog.html>.

¹² <http://gnis-ld.org/resource/gnis-sameas-geonames.ttl>.

¹³ <https://github.com/blake-regalia/usgs-triplifier/tree/master/lib/align>.

4 The Dataset

The GNIS and other USGS products are public domain datasets¹⁴ that are maintained, updated, and supported by the U.S. Federal Government. We created the GNIS-LD as a 5-star linked open dataset version of the GNIS for USGS to maintain. The GNIS dataset as of February 1, 2018 contains over 2.27 million features for the United States (see Table 1a) together with their geometries, alternative names, types, containment relations, elevations, historic notes, and so forth. It contains man-made features such as cities as well as natural features such as mountain peaks and ranges across different scales from single buildings to entire states. Our Linked Data triplication process yields 37 million triples for the GNIS dataset alone. These features are made up of 66 distinct types, with the top 10 feature types shown in Table 1b.

It is worth putting the GNIS-LD into context by describing its relation to [GeoNames.org](https://www.geonames.org/)¹⁵ and [LinkedGeoData](https://linkedgeo.org/)¹⁶. Most importantly, these two resources either directly imported or indirectly inherited a significant portion of their US data from the GNIS at one point in time. However, they do not necessarily reflect the current version of the GNIS and also allow for volunteered contributions from the community. GNIS-LD is an authoritative, comprehensive, triplified version of the most up-to-date dataset for the names of places in the US. Furthermore, whereas GeoNames is not 5-star Linked Data and has no SPARQL endpoint, and LinkedGeoData supports only a subset¹⁷ of the SQL MM spatial specification (via non-standard Extensible Value Testing filter functions under the `bif:` prefix in SPARQL), GNIS-LD offers a 5-star Linked Dataset with full GeoSPARQL support¹⁸. Finally, our dataset is designed to be compatible with high-resolution, complex geometries provided by USGS. We show some preliminary work integrating one of these datasets with the GNIS-LD in Fig. 1. In this example, the `nhd:gnisFeature` predicate links the sole geometry of Lake Tahoe to its GNIS feature which represents the naming of the water body.

The GNIS gazetteer is particularly important as it acts as a nexus between other datasets and to support interaction and workflows of human users (as compared to software agents) which is most often done using place names instead of geometries. For instance, and as depicted in Fig. 2, a USGS station from the WaterWatch program is located inside/at a segment of Tobesofkee Creek near the city of Macon, GA; thereby linking measurement results to the creek and city. As the city record from GNIS is linked to DBpedia via an `owl:sameAs` relation, one can get additional information e.g., demographic data, about the city.

¹⁴ https://www2.usgs.gov/laws/info_policies.html#copyrights.

¹⁵ <https://old.datahub.io/dataset/geonames-semantic-web>.

¹⁶ <http://linkedgeo.org/About>.

¹⁷ <http://docs.openlinksw.com/virtuoso/sqlrefgeospatial/>.

¹⁸ <https://wiki.apache.org/marmotta/GSoC/2015/MARMOTTA-584>.


```
# see <http://phuzzy.link/context/gnis-ld> for prefix mappings

# how many lakes are there in California? result: 73,497
select (count(?lake) as ?lakes) {
  ?lake a nhd:Reservoir ;
  gnis:state gnisf:California .
}

# what is the total length of the Truckee River? result: 148.3 km
select (sum(?segment_length) as ?river_length) {
  ?river nhd:gnisFeature/rdfs:label "Truckee River"@en ;
  ago:geometry ?geometryUri .

  ?geometryUri ago:boundingBox ?boundingBox ;
  ago:length [
    qudt:numericValue ?segment_length ;
    qudt:unit unit:M ;
  ] .
}
```

Listing 3 Showcasing a few sample queries on the GNIS-LD and NHD geometries.

The screenshot shows the USGS Linked Data interface for the resource `nhd:44560536`. The interface is split into two main sections: a metadata panel on the left and a map on the right.

Metadata Panel (Left):

- URI:** `nhd:44560536`
- rdftype:** `nhd:Average_Water_Elevation_Perennial_Lake_or_Pond`
- ago:geometry:**
 - USGS URL: `usgs-mt/foolypor/nhd/44560536`
 - Well Known Binary (WKB)
 - Well Known Text (WKT)


```
MULTIPOLYGON((-119.989951524192 39.248762057447,-119.989936124192 39.248754257447,-119.98996124192 39.2487774724111,-119.989934124192 39.248773207447,-119.989929390899 39.248790027447,-119.989937390899 39.248742724114,-119.988851924193 39.248737807447,-119.988863190899 39.2487412724114,-119.988848324194 39.248737807447,-119.988249908999 39.2487310724114,-119.987800790899 39.248664057449,-119.987353360899 39.2486404397449,-119.987281508999 39.248608697449,-119.988721390899 39.248566874117,-119.986913908999 39.2485524724117,-119.986705990899 39.2485290724118,-119.986618291099 39.248500747417,-119.9866357941099 39.248503294419))
```
 - GeoJSON
- nhd:reachCode:** `*16050101000339*`
- nhd:elevation:** (empty)
- nhd:gnisFeature:** `Lake Tahoe`
- nhd:arba:** (empty)
- nhd:resolution:** `nhd:High_Resolution`
- usgs:modified:** `March 11, 2012`
- usgs:permanentIdentifier:** `*44560536*`

Map (Right): A satellite-style map showing the geographical location of Lake Tahoe, outlined in blue. The map includes a zoom control (+/-) in the top right corner.

Fig. 1. A geographic feature with polygon geometry converted to linked data and linked to its GNIS record, as displayed in our web interface.

5 User Interface

When it comes to choosing a Linked Data front-end interface that supports the display of and interaction with geospatial data, one can select from a small number of existing solutions. GeoLink [11, 14], Sextant [4, 15], and SPEX [20] each take a unique approach to exploring geographic data, which can have many possible modes of interaction depending on the nature of the dataset, e.g., trajectories, time series, complex geometries, and so on, as well as browsing paradigms, i.e., whether to use an interactive map, faceted browser, graph-view, or something in between. Other, non geo-specific, approaches focus on modularity. Among these, Linked Data Reactor [10], Uduvudu [13], LodLive [5], and Fresnel [17] unite under the common goal of building Linked Data interfaces out of reusable components.

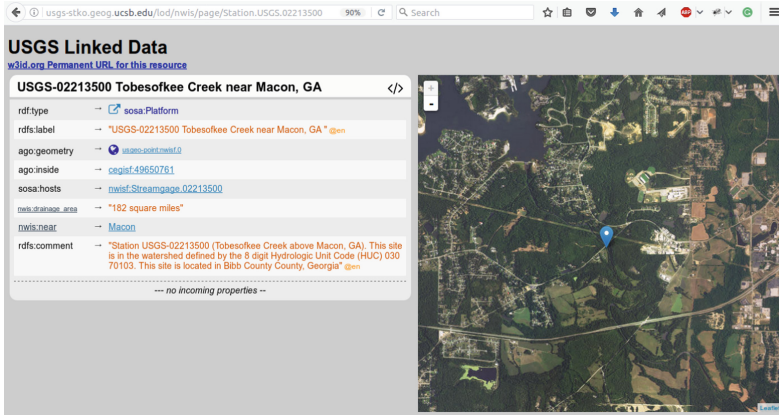


Fig. 2. A streamgage measurement station in the Tobesofkee Creek near Macon, CA annotated using the SOSA/SSN ontology.

For GNIS-LD we decided to combine both approaches by maximizing reusability and at the same time offering support for geographic data beyond points. The resulting interface named Phuzzy.link [18], is similar to Pubby¹⁹ insofar that it describes each resource by showing its outgoing properties in a tabular format with hyperlinks for locally dereferenceable URIs and special formatting for certain datatyped literals (e.g., `xsd:date` values). Where our approach differs from previous works is how components are sourced and how the content-agnostic interfaces is generated. Our interface queries the SPARQL endpoint directly from the client and creates human-readable representations of the resource using a customizable configuration that is tailored to each dataset, either by the provider, the community, or both. To keep displays between pages consistent and readable, rows are displayed in order according to the priority assigned to each predicate by the data provider. For example, `rdf:type` is among our highest priority for outgoing predicates, so it will be displayed as the first row for each resource that has an `rdf:type` triple, followed by its `rdfs:label`, and so forth.

The text for hyperlinks that point to adjacent resources will also be substituted by their `rdfs:label` if one was returned in the initialization SPARQL query used by the interface. For incoming triples, the interface also asks for a subject's `rdf:type` if it is available so that the interface can organize the results into collapsible groups as shown in Fig. 3. This helps reduce the clutter on the screen for common objects that are linked to by many triples, such as counties and states.

We designed the interface to embed special interactive features for select resource types. Namely, we support: unit conversion for quantities such as elevation values, display format toggling for date and time literals, interactive map

¹⁹ <http://wifo5-03.informatik.uni-mannheim.de/pubby/>.

Santa Barbara County	
rdfs:type	→ gnis:County
rdfs:label	→ "Santa Barbara County"@en
gnis:featureId	→ "277306"
gnis:countyCode	→ "083"
gnis:countyName	→ "Santa Barbara"@en
gnis:state	→ State of California
gnis:country	→ gnis:United_States

is gnis:county	of
	1 Arch
	7 Bars
	5 Basins
	32 Bays
	2 Beaches
	East Beach
	Leadbetter Beach
	4 Benches
	1 Building
	52 Capes

Fig. 3. Santa Barbara county is the `gnis:county` of many GNIS features, which are grouped and collapsed together by their `rdfs:type` in the incoming properties section.

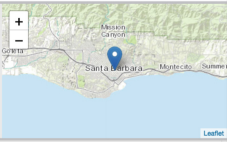
Santa Barbara	
Bound on the north by the Santa Ynez Mountains, on the east by Montecito, on the south by the Santa Barbara Channel, and on the west by Goleta. (US-T121)	
	
rdfs:type	→ gnis:PopulatedPlace
rdfs:label	→ "Santa Barbara"@en
geo:geometry	→ urigo:point:gnis:1861401
gnis:alternativeName	→ "La Laguna de la Concepcion"@en
gnis:citation	→ "U.S. Geological Survey. Geographic Names Phase 1 data compilation (1976-1981). 31-Dec-1981. Priority from U.S. Geological Survey 1:24,000-scale topographic maps (or 1:50K Puerto Rico 1:20K) and from U.S. Board on Geographic Names files. In some instances, from 1:62,500 scale or 1:62,500 scale maps."@en
gnis:county	→ gnis:state:California.Santa_Barbora
gnis:dateFeatureCreated	→ January 16, 1981 ¹ xsd:date
gnis:dateFeatureEdited	→ June 14, 2015 ¹ xsd:date
gnis:dateNameCreated	→ January 16, 1981 ¹ xsd:date
gnis:description	→ "Bound on the north by the Santa Ynez Mountains, on the east by Montecito, on the south by the Santa Barbara Channel, and on the west by Goleta. (US-T121)"@en
gnis:elevation	→ gnis:Elevation:498
gnis:featureId	→ "1861401"

Fig. 4. The interface showing GNIS data about Santa Barbara, CA including its location on a map, available at <https://bit.ly/2DPZGM4>.

plotting for places with geometries, and the option to download a feature's source geometry data in a variety of serialization formats. With the exception of the last feature, all interactivity is handled in-browser by the client so that the endpoint's resources can be reserved for executing SPARQL queries. We discuss these features in greater detail below.

GNIS survey data for elevation above sea level are recorded in imperial units (ft). Since many users will encounter the need to convert these quantities to meters or kilometers, we approached unit conversion as the need for a modular feature within the user interface that can be adapted to any quantity types. By utilizing the QUDT ontology²⁰, we preemptively download conversion rates to a quantity's possible units given in the QUDT vocabularies. A user can then select from a dropdown menu of available units to convert a quantity entirely in-browser, i.e., without additional queries to the server.

To make geometry data available from within the user interface, an interactive element can be expanded by clicking the globe icon that appears next to a geometry's URI, shown in Fig. 1. From there, a list of possible serialization formats is shown with download options next to each item. Clicking the option to display the geometry as text or to download it as a file both trigger an asynchronous HTTP request set with the appropriate 'Accept' headers.

Users who explore Linked Data through a front-end are not always interested in high-level views that encapsulate the underlying RDF. For those who want to see how an ontology is being utilized or to simply access a resource's RDF

²⁰ <http://www.qudt.org/release2/qudt-catalog.html>.

closure without writing a SPARQL query, we provide a display toggle (</>) that shows the RDF for the current resource’s outgoing triples in a textbox of syntax-highlighted Turtle.

6 Availability and Sustainability

The GNIS-LD and future Linked Data versions of USGS datasets are made permanently and openly available as a public data service²¹. The repository can be queried via a public SPARQL endpoint at <http://gnis-ld.org/sparql/select>; see also <http://yasgui.org/short/H130H1XcM>. All IRIs for features and geometries as indicated by the prefixes `nhd`, `nhdh`, `usgs`, `gnis`, `gnisf`, `usgeo-point`, `usgeo-polygon` and so forth, support content-negotiation for RDF or geometry data and can be dereferenced in a web browser to access the human-readable representations in our interactive user interface.

Our dataset is also made available on datahub.io²² as part of the US Geological Survey organization. The datahub.io entry includes references for:

- **VoID description**—Machine readable metadata about the dataset.
- **GNIS feature definitions**—Feature type vocabulary for GNIS.
- **GNIS-LD RDF dump**—The entire GNIS dataset as RDF.
- **USGS-LD SPARQL endpoint**—The SPARQL endpoint for live data.
- **USGS-LD SPARQL service description**—Machine readable metadata about the SPARQL endpoint.

Updates to the underlying source data will subsequently trigger updates to the endpoint’s triple store and RDF data dumps.

7 Summary and Future Work

In this resource paper, we presented an authoritative Linked Dataset for the Geographic Names Information (GNIS) System that complements existing crowd-sourced and non-authoritative resources. The datasource contains millions of places in the United States together with their geometries, alternative names, types, containment relations, elevations, historic notes, and so forth. The data contains places across more than 60 feature types and across different scales ranking from places of worships to rivers. Accompanying the dataset, we also provide an ontology, a SPARQL endpoint, metadata about the dataset and endpoint, RDF data dumps, and a dereferencing web interface with content negotiation for RDF and geometry data. Co-reference resolution links to GeoNames and DBpedia are provided as `owl:sameAs` relations. GNIS-LD is a milestone for the linked geodata community as it is among the first and few authoritative geographic datasets released in direct collaboration with the US government

²¹ <https://www2.usgs.gov/publishing/policies.html>.

²² <https://old.datahub.io/dataset/geographic-names-information-system-gnis>.

agencies that created and maintained these data, and it is important for the Semantic Web because places in general act as nexuses that connect actors, events, and objects.

We presented preliminary work for how this resource aligns with upcoming datasets such as the DLGs²³ and National Map data more broadly as well as with other authoritative data sources such as USGS WaterWatch sensor data. In the future, we will aim at providing further links to other Linked Data sources such as Getty's TGN as well as integration with other types of sensor data.

References

1. Atemezing, G.A., Abadie, N., Troncy, R., Bucher, B.: Publishing reference geodata on the web: opportunities and challenges for IGN France. In: TC/SSN at ISWC, pp. 9–20 (2014)
2. Atemezing, G.A., Troncy, R.: Comparing vocabularies for representing geographical features and their geometry. In: Terra Cognita 2012 Workshop, vol. 3 (2012)
3. Bereta, K., Koubarakis, M.: Ontop of geospatial databases. In: Groth, P., Simperl, E., Gray, A., Sabou, M., Krötzsch, M., Lecue, F., Flöck, F., Gil, Y. (eds.) ISWC 2016. LNCS, vol. 9981, pp. 37–52. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46523-4_3
4. Bereta, K., Nikolaou, C., Karpathiotakis, M., Kyzirakos, K., Koubarakis, M.: Sextant: visualizing time-evolving linked geospatial data. In: Proceedings of the 2013th International Conference on Posters and Demonstrations Track, vol. 1035, pp. 177–180. CEUR-WS. org (2013)
5. Camarda, D.V., Mazzini, S., Antonuccio, A.: LodLive, exploring the web of data. In: Proceedings of the 8th International Conference on Semantic Systems, pp. 197–200. ACM (2012)
6. Varanka, D.E., Wendel, J., Janowicz, K.: Technology transfer to incorporate linked geospatial data for national mapping (under review)
7. Debruyne, C., Clinton, É., McNerney, L., Nautiyal, A., O'Sullivan, D.: Serving Ireland's geospatial information as linked data. In: International Semantic Web Conference (Posters and Demos) (2016)
8. Guptill, S.C.: An enhanced digital line graph design. Technical report, USGPO; Books and Open-File Reports Section, US Geological Survey (1990)
9. Hogan, A., Hitzler, P., Janowicz, K.: Linked dataset description papers at the semantic web journal: a critical assessment. *Semant. Web* **7**(2), 105–116 (2016)
10. Khalili, A., de Graaf, K.A.: Linked data reactor: towards data-aware user interfaces. In: Proceedings of the 13th International Conference on Semantic Systems, SEMANTICS (2017)
11. Krisnadhi, A., et al.: The GeoLink modular oceanography ontology. In: Arenas, M., et al. (eds.) ISWC 2015. LNCS, vol. 9367, pp. 301–309. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25010-6_19
12. Kuhn, W., Kauppinen, T., Janowicz, K.: Linked data - a paradigm shift for geographic information science. In: Duckham, M., Pebesma, E., Stewart, K., Frank, A.U. (eds.) GIScience 2014. LNCS, vol. 8728, pp. 173–186. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11593-1_12

²³ <https://lta.cr.usgs.gov/DLGs>.

13. Luggen, M., Gschwend, A., Anrig, B., Cudré-Mauroux, P.: Uduvudu: a graph-aware and adaptive UI engine for linked data. In: LDOW@ WWW (2015)
14. Mai, G., Janowicz, K., Hu, Y., McKenzie, G.: A linked data driven visual interface for the multi-perspective exploration of data across repositories. In: VOILA@ ISWC, pp. 93–101 (2016)
15. Nikolaou, C., Dogani, K., Kyzirakos, K., Koubarakis, M.: Sextant: browsing and mapping the ocean of linked geospatial data. In: Cimiano, P., Fernández, M., Lopez, V., Schlobach, S., Völker, J. (eds.) ESWC 2013. LNCS, vol. 7955, pp. 209–213. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41242-4_26
16. Perry, M., Herring, J.: OGC GeoSPARQL-A geographic query language for RDF data. OGC Implementation Standard, September 2012
17. Pietriga, E., Bizer, C., Karger, D., Lee, R.: Fresnel: a browser-independent presentation vocabulary for RDF. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 158–171. Springer, Heidelberg (2006). https://doi.org/10.1007/11926078_12
18. Regalia, B., Janowicz, K., Mai, G.: Phuzzy.link: A SPARQL-powered client-sided extensible semantic web browser
19. Regalia, B., Janowicz, K., McKenzie, G.: Revisiting the representation of and need for raw geometries on the linked data web. In: Proceedings of the Workshop on Linked Data on the Web (LDOW), CEUR Workshop Proceedings (2017)
20. Scheider, S., Degbelo, A., Lemmens, R., van Elzakker, C., Zimmerhof, P., Kostic, N., Jones, J., Banhatti, G.: Exploratory querying of SPARQL endpoints in space and time. *Semant. Web* **8**(1), 65–86 (2017)
21. Zhu, R., Hu, Y., Janowicz, K., McKenzie, G.: Spatial signatures for geographic feature types: Examining gazetteer ontologies using spatial statistics. *Trans. GIS* **20**(3), 333–355 (2016)