## Geospatial Semantics: Georgia O'Keeffe in Context

# Monica Hinga and Yingfan Wang 7 May 2021

## 1 Introduction

First described in 2001 by Tim Berners-Lee, the semantic web is as much a vision of a better way to access information as it is a set of standards and tools for web developers to use to link information across the World Wide Web [1]. Although the possibilities of semantic web agents excited many people back then, and continue to do so today, the World Wide Web of the future envisioned by Berners-Lee has not come to fruition. Even in 2006, Berners-Lee himself and his collaborators were expressing just how much more difficult creating a linked World Wide Web really is compared to the straightforward description from only a few years earlier [2]. By 2006 there had been efforts to develop standards, including RDF triple-stores and OWL, that are still in use today. Using these standards and encouraging web developers and tech companies to link their data, however, has been more easily said than done.

The challenges of adopting and using these standards include such things as the difficulty in applying very formal ontologies to web content, and the challenges in getting access to the kinds of graphs developed by Google and its ilk. The utility of the semantic web, however, cannot be ignored. Museums, in particular, have much to gain from a linked World Wide Web because their mission is educational. Since the early-2000s, there have been attempts by museums to use the semantic web in order to reach their audiences. MuseumFinland, for example, was intended to link artifacts with the materials from which they were made, their makers, and their collections [3]. Although the Finnish effort won an award in 2004, multiple Internet searches, albeit in English, were unsuccessful in finding any evidence of this project being in use in 2021. That semantic web projects for museums happen in fits and starts seems to be well-recognized by the museum community. Semantics are important to how museums are organized and how they present information in a way that is accessible to their patrons, but the semantic web is not well-understood or utilized by museums[4].

The Georgia O'Keeffe Museum has been one of the few museums to utilize the semantic web for their virtual collection. They have already been able to connect their rich collection of Georgia O'Keeffe's works, writings, and personal belongings to her place in American art as well as with the many people who were part of her life [5]. The work that has already been done at the Georgia O'Keeffe Museum to create a vibrant and informative experience for visitors to their virtual collection can be expanded to even further enhance website visitors' ability to experience Georgia O'Keeffe.

In this document, we:

• propose a project to further enhance the Georgia O'Keeffe Museum's virtual collection using the semantic web,

- identify tools that may be used to achieve the goals described in the project,
- make a recommendation to the Georgia O'Keeffe Museum for tools to be adopted,
- and provide semantic web queries that could be used by the Georgia O'Keeffe Museum for their virtual collection.

## 2 Project Proposal

## 2.1 Georgia O'Keeffe

One of the most important American artists of the 20th century, Georgia O'Keeffe is known for her unique way of seeing nature. Some of her most important early works were her paintings of flowers. Later, she moved to New Mexico where she documented the American Southwest through her unique vision [5].

Although some Americans may be familiar with the striking landscapes of her New Mexican home, those who are not familiar with the American Southwest may gain a deeper understanding of O'Keeffe's art by understanding the geography of her paintings.

## 2.2 Proposal: Geospatial Ontologies to Enrich the Georgia O'Keeffe Museum Online Resources

The primary goal of this project is to enhance the museum's virtual collection by providing additional contextual information for artworks. As this would be a large undertaking, we will focus on providing geographical context for O'Keeffe's New Mexico paintings. Although this omits her works depicting other themes such as architecure and her famous paintings of flowers, a similar model could be applied to artworks with other themes in the future. Our work will have three phases:

First, geographical and geospatial ontologies will be evaluated for suitability. At this point two ontologies have been identified for evaluation, but there may be more. At a minimum, the USGS Geospatial Semantics and Ontology and GeoNames [6, 7] will be evaluated. The suitability of available geospatial ontologies will be evaluated based on such diverse criteria as licensing, relevance of information to the Georgia O'Keeffe Museum, and ease-of-use.

Once an ontology, or ontologies, has been identified, then the process of developing SPARQL queries for the Georgia O'Keeffe Museum will begin. SPARQL queries will be developed for a sample of O'Keeffe's New Mexico works. For example, Black Mesa Landscape, New Mexico / Out Back of Marie's II depicts a stylized vision of the area around Black Mesa. It may be helpful for someone who is interested in O'Keeffe's work to have a better understanding of where Black Mesa is, what it looks like, and some information about the geology.

Finally, if the results are satisfactory, we will develop a proposed workflow for incorporating geospatial information into the already rich virtual collection on the Georgia O'Keeffe Museum website.

## 3 Summary of Findings

Before getting into the work of identifying an ontology, or ontologies, suitable for use at the Georgia O'Keeffe Museum, it is important to take the time to discuss the meanings of "geographical" and "geospatial." "Geospatial" encompasses both features and geometries, where a "feature" is any place in the real world, e.g. a mountain, a park, or a shopping mall and a "geometry" is any geometric shape, including a single pixel, a circle, or polygon [8]. Interestingly, geospatial data is a superset of geographical data. "Geographical" data is only the coordinate system, i.e. the geometry part of geospatial data.

It is possible to use geospatial ontologies alongside other ontologies to provide context to O'Keeffe's work.

## 3.1 Methodology

Our project was completed in three phases: identify existing geospatial and geographic ontologies, evaluate them for suitability, and once an ontology, or ontologies, has been selected, write queries for a sample of O'Keeffe's New Mexico paintings.

#### 3.1.1 Identify Existing Ontologies

Because geographical and geospatial data are useful for everything from searching for a sushi restaurant, to getting directions to a party, to scientific and military applications, there are multiple ontologies available for geospatial data. It should be noted that although making geospatial data available on the semantic web is relatively easy, developing a query language that can handle geospatial data appropriately is challenging and there are competing standards for this [8]. For this reason, the issue of query languages is also discussed briefly where appropriate. Note that the Georgia O'Keeffe Museum already uses SPARQL, so compatibility with SPARQL is an important factor in selecting the best option for the Georgia O'Keeffe Museum.

GeoSPARQL is, as the name implies, a query language for geospatial data based on SPARQL as well as a standard for the representation of geospatial data. It was developed and is maintained by the Open Geospatial Consortium (OGC), a standards organization made up of members from government, academia, and industry [8]. As with many other areas of computing, there are multiple standards for querying geospatial data, each with its own raison d'être. In the case of GeoSPARQL, it seeks to better handle queries for information that is not explicitly represented in available data[8]. For example, queries like "what lakes are within 10 miles of Minneapolis?" rely on how the data are represented. If the data do not reference Minneapolis, then SPARQL cannot handle this query.

The GeoSPARQL standard leverages some of the advantages of RDF, such as its ability to better handle hierarchical information, combined with existing spatial databases, to create one of the most commonly used standards for querying geospatial data on the semantic web [8]. Although it is commonly used, there are almost no complete implementations of this standard. The most popular implementations, none of which are compelete, are Apache Marmotta, Apache Jena, and Parliament. Unfortunately, even among these more popular projects, there has been little development. In the case of Apache Marmotta, which was only available for PostgreSQL, the project was retired in 2020 [9]. Parliament is still being developed, but it also is only available for PostgreSQL and is poorly documented[10]. Of

the available GeoSPARQL implementations listed here, Apache Jena seems to be the most popular, has the best documentation, and permits the use of SPARQL queries[11]. Of course, Jena is not without challenges. It is Java-based, which means that the Georgia O'Keeffe Museum would have a learning curve for using it.

W3C Basic Geo Vocabulary In 2003, the W3C Semantic Web Interest Group created their Basic Geo Vocabulary, which was later incorporated into a project led by the W3C Geospatial Incubator Group [12, 8]. Their ontology is intended for use with RDF, but can also be used in standard XML documents. Although it started strong and there are other ontologies written in compliance with this standard, W3C never fully adopted it and it never got out of the incubator phase [12, 8]. This project is not currently active and has not been updated in more than ten years [12].

GeoNames GeoNames is an ontology for geospatial data, including toponyms, or place names, and altitudes. It has been used by Apple in Mac OS X Snow Leopard, by Canonical in Ubuntu Linux, Etsy, and FourSquare among others. It provides downloadable .csv files, available for free under a Creative Commons License [6]. Though popular, it does not have a SPARQL endpoint. Instead, getting a data dump from GeoNames or using the search tool they refer to as "Mother Earth" are preferred methods [6, 13].

USGS Geospatial Semantics and Ontology The US Geological Survey (USGS) is one of the member organizations of the OGC, the organization that maintains the GeoSPARQL standard, but it also does its own work on linked geospatial data and the uses of the semantic web for understanding topographical and geospatial information [7]. Although they have studied the usefulness of the semantic web for geospatial applications, their ontology has been included in OGC's GeoSPARQL [7, 14].

USGS has, however, made strides towards using linked data to improve the availability of geospatial data. Their methods may be useful to the Georgia O'Keeffe Museum at a later phase of their semantic web project, but this work is out of scope of this project [15].

#### 3.1.2 Evaluate Ontologies

Each of them were evaluated based on the following criteria: relevance of available data to the Georgia O'Keeffe Museum, licensing, ease-of-use, and compatibility with SPARQL queries, which are already being used by the Georgia O'Keeffe Museum.

#### GeoSPARQL

- Data: The data that can be accessed using GeoSPARQL is relevant to the Georgia O'Keeffe Museum. Many of the place names in Georgia O'Keeffe's body of work are vague. Even when limited the area to northern New Mexico, it can be unclear, especially for people who have never visited New Mexico, which locations are referenced in the titles of O'Keeffe's works. Because GeoSPARQL queries are very good at handling specific coordinates and the Georgia O'Keeffe Museum has already gone through the process of identifying coordinates for O'Keeffe's body of work, GeoSPARQL queries could be useful to them.
- Licensing: The OGC has elected to use its own license for GeoSPARQL. The OGC license is flexible and similar to other popular open source licenses [16]. One concern

with this license is that there have been some problems when trying to use it with DBpedia Archivo, which is a popular ontology archive [17]. It should be noted that this is still an open issue on their issue tracker and it appears that they have plans to resolve it, but in the immediate future compatibility with DBpedia is compromised.

- Ease-of-Use: As mentioned above, there are numerous implementations of the GeoSPARQL standard. Of these implementations, Apache Jena is the best choice for the Georgia O'Keeffe Museum because it also handles standard SPARQL queries.
- Compatibility with SPARQL: GeoSPARQL and SPARQL are meant to work together. GeoSPARQL only provides additional capabilities for queries that are not handled well by standard SPARQL. Apache Jena handles both SPARQL and GeoSPARQL queries. The downside for the Georgia O'Keeffe Museum of Jena is that it is Javabased. Because of this, it has the capacity to be more versatile and explains some of its popularity, but the Georgia O'Keeffe Museum does not currently use anything like Jena and, depending on future plans, it may be overkill.

### W3C Basic Geo Vocabulary

- Data: That data that can be queried using this standard is similar to that which can be queried using GeoSPARQL.
- Licensing: W3C has its own license language which is similar to more widely used open source licenses [12]. There are no issues with it being recognized by other systems as with GeoSPARQL.
- Ease-of-Use: Although there are existing implementations of this standard and it can be used with RDF and XML, it is not being actively worked on by W3C. Orphaned standards provide additional challenges over standards that are actively being maintained.
- Compatibility with SPARQL: SPARQL can handle queries of W3C Basic Geo Vocabulary, so it would be easy for the Georgia O'Keeffe Museum to implement.

## GeoNames

- **Data:** GeoNames is a large dataset of placenames and locations. It is one of the most popular geospatial ontologies.
- Licensing: GeoNames is made available under the Creative Commons 4.0 license. This is a standard license that permits sharing and making changes to GeoNames data as long as it is attributed to GeoNames [6].
- Ease-of-Use: GeoNames can be used by downloading the data and querying the downloaded copy using a database of the user's choice. It can also be accessed through their "Mother Earth" website [6]. Although this provides additional flexibility to developers, it is not ideal for semantic web applications. This may explain why it has been used in operating systems and for websites run by large tech companies.
- Compatibility with SPARQL: As stated previously, GeoNames does not have a SPARQL endpoint [6].

USGS Geospatial Semantics and Ontology Because USGS has been a contributor to the OGC for GeoSPARQL, for the purposes of this project they can be considered a subset of GeoSPARQL. As such, the analysis for GeoSPARQL above applies to the USGS project. They are mentioned separately here only because they have done some interested work on using linked data for geospatial applications that, while out of scope for this project, may be interesting to the Georgia O'Keeffe Museum should they decide to emphasize geospatial information in their virtual collection [7].

**Recommendation** We recommend that the Georgia O'Keeffe Museum adopt GeoSPARQL as implemented in Apache Jena for geospatial queries. Although Apache Jena is Java-based and would be a departure for the Georgia O'Keeffe Museum from their current systems, it is more flexible and better maintained than other implementations.

We tested Jena as well as quieries written using the W3C standard and found that documentation for Jena was easier for new users. High quality documentation mitigates the challenges of moving towards a Java-based standard.

#### 3.1.3 Write SPARQL Queries

**Using Jena** Jena is a Java API that is used to support semantic web applications. Jena uses an interface called a Model to contain RDF triple stores. Without getting into specifics, it should be noted that the ModelFactory class provides a more accessible way to create RDF models.

In addition to Jena, we also conducted searches using a more general-purpose semantic web data source: Wikidata. While we had intended Wikidata to merely serve as a point of comparison, we ultimately found that Jena is more useful for custom datasets, but that Wikidata can be leveraged to reduce the burden of data collection. In fact, it would be advisable that the Georgia O'Keeffe Museum consider using both standards together. Below, we see examples queries in 5 groups. Groups 1-4 demonstrate searches using Jena and group 5 is an example Wikidata search.

Sample Queries For this project, we focused on writing queries for Georgia O'Keeffe's paintings in the Georgia O'Keeffe Museum's collection depicting the New Mexico landscape. We have identified several classes of queries that demonstrate the capabilities of Apache Jena GeoSPARQL and how it may be useful for the Georgia O'Keeffe Museum. Imaging that you are a visitor to the Georgia O'Keeffe Museum's virtual collection who wants to learn more about Georgia O'Keeffe's painting Untitled / Red and Yellow Cliffs from 1940 and seen in 1. The cliffs in the painting are similar to many other cliffs in the American Southwest. The following sample queries demonstrate how Apache Jena can be used to learn more about the geographical features in the painting and the painting itself.

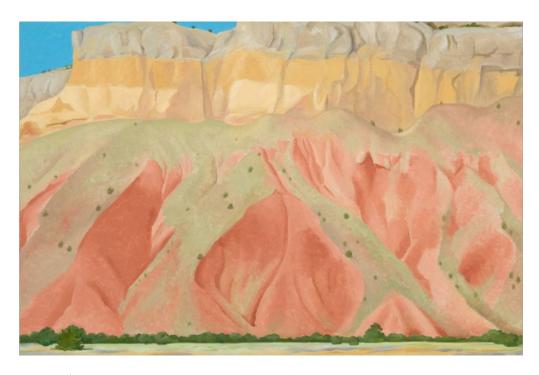


Figure 1:  $Untitled / Red \ and \ Yellow \ Cliffs$ , 1940 by Georgia O'Keeffe from the Georgia O'Keeffe Museum's collection. Many of the example searches relate to this painting.

## RDF Data from XML File on Georgia O'Keeffe Museum Website

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:cd="http://www.recshop.fake/cd#">
<rdf:Description
rdf:about="https://collections.okeeffemuseum.org/object/84/">
<cd:code>CR0998</cd:code>
<cd:code>CR0998</cd:code>
<cd:name>Untitled (Red and Yellow Cliffs)</cd:name>
<cd:latitude>36.335363</cd:latitude>
<cd:longitude>-106.495020999999</cd:longitude>
<cd:population>1058</cd:population>
</rdf:Description>
</rdf:RDF ></rdf:RDF</pre>
```

The above RDF data is taken from the Georgia O'Keeffe Museum website. It is typical of RDF data available for other paintings in their collection. Notice that the we see not only latitude and longitude but because the name of the painting also includes a feature, i.e. "cliffs", this is an example of a work that is suited to geospatial searches.

Group 1: Search one property of one entity For this group of searches, we know the link of the painting in which we are interested, in this case, the painting shown above in

Figure 2: Here we see example output from group 1, the output from searching one property of one entity.

Figure 3: In this image, we see an example result for group 2. Instead of returning a data element for only one painting, we return an element for multiple works in the Georgia O'Keeffe Museum collection.

Fig. 1. Here we know the link where the painting is available. We can use that to find the code associated with the painting.

```
prefix cd: <http://www.recshop.fake/cd#>
SELECT ?code
WHERE {<https://collections.okeeffemuseum.org/object/84/> cd:code ?code}
```

In this example, we return the value assigned to the painting for code but a similar search could be used to return any of the data elements in the painting, including the name of the painting or its latitude.

Group 2: Search one property of all entities — It is also possible to return one property from all entities, e.g., the names of all paintings or all of their urls. Although this search may not be as useful for user facing applications, it could be used internally to make management of the Georgia O'Keeffe Museum's data easier. It might also be useful for researches who want information about every item in the collection. We see the results of this type of search in Fig. 3

```
prefix cd: <http://www.recshop.fake/cd#>
SELECT ?x ?name
WHERE {?x cd:name ?name}
```

Figure 4: This is a sample result from a search from group 3. We see the latitude and longitude from Fig. 1

Group 3: Search two or more properties of one entity Now we consider the case in which someone wishes to return more than one data element for one item in the collection. We anticipate this to be one of the more useful types of searches, especially for geospatial information because at least two data elements would be needed: latitude and longitude. (See Fig. 4)

```
prefix cd: <http://www.recshop.fake/cd#>
   SELECT ?latitude ?longitude
   WHERE {{ <https://collections.okeeffemuseum.org/object/84/> cd:latitude
?latitude}   UNION   {<https://collections.okeeffemuseum.org/object/84/> cd:longitude
?longitude}}
```

Group 4: Search for artworks with given geospatial features In this group, we demonstrate how to handle searches for paintings depicting the same feature. It should be noted that this relies on using the titles of works to conduct the search, but this functionality could be easily expanded by adding additional data elements. If one wanted to search for all paintings of cliffs, one could search titles of works for the word "cliff." (Results are seen in Fig. 5.) This type of search may also be useful for users who want to find a painting that they liked but of which they could not remember the title, e.g. paintings of skulls or zinnias.

Group 5: Searches using Wikidata Apache Jena is very good for searching custom data. If you need to include your own dataset on the semantic web, tools like Jena provide the amount of flexibility that most developers would want in order to produce a high quality user experience. There are, however, preexisting datasets that may contain relevant information. Given that semantic web data sources become more and more rich over time, the possibilities here reach farther than geospatial data. Within the scope of this project, however, we suggest that Wikidata could be used to search for geospatial features because

Figure 5: For users interested in all paintings with titles containing the word "cliff," this is the result of that search.

the amount of information already present on Wikidata would mean less duplication of efforts in the seemingly simple, but in actuality very expensive and time-consuming, task of data collection. Here we demonstrate a search for the latitude and longitute of the Georgia O'Keeffe Museum using Wikidata.

Another advantage of Wikidata that may make it useful should the Georgia O'Keeffe Museum want to attract international visitors to its virtual collection is that it can search data in multiple languages. Full language localization is expensive, but Wikidata already has some information translated into other languages. Taking advantage of localization work that has already been done can bring such projects closer to being within reach.

In order to find out how much information on Wikidata would be relevant for the Georgia O'Keeffe Museum we also conducted searches of Georgia O'Keeffe's works. We did not find entries for her paintings. As such, we recommend that if the Georgia O'Keeffe Museum chooses to use Wikidata for geospatial searches, it does not rely on it solely.

```
Search data from WIKIDATA by SPARQL
For example, search six places within 1km of the Georgia O'Keeffe Museum
PREFIX wd: <a href="http://www.wikidata.org/entity/">http://www.wikidata.org/entity/>
PREFIX wdt: <a href="http://www.wikidata.org/prop/direct/">http://www.wikidata.org/prop/direct/</a>
SELECT ?place ?placeLabel ?location ?instanceLabel
WHERE
  wd:Q1509304 wdt:P625 ?loc .
  SERVICE wikibase: around {
       ?place wdt:P625 ?location .
       bd:serviceParam wikibase:center ?loc .
       bd:serviceParam wikibase:radius "1" .
  OPTIONAL {
                  ?place wdt:P31 ?instance }
  SERVICE wikibase: label { bd:serviceParam wikibase:language "en" }
  BIND(geof:distance(?loc, ?location) as ?dist)
} ORDER BY ?dist
Limit 6
```



Figure 6: Wikidata is a rich source of data that has already been collected. Here we see the output for a search of the latitude and longitude of the Georgia O'Keeffe Museum.

## 4 Conclusions and Future Work

The possibilities for the Georgia O'Keeffe Museum to leverage the semantic web to provide geospatial context to their virtual collection are real and attainable. They already have location information for many of her paintings and there are open source, easy-to-use tools available that the Georgia O'Keeffe Museum can leverage. That is not to say that this work would be without challenges. Apache Jena is Java-based which means a steeper learning curve and there would be some need to discern exactly how best to use the geospatial information available to the Georgia O'Keeffe Museum.

As the Georgia O'Keeffe Museum considers its semantic web future, geospatial queries should not be thought of only in the context of Georgia O'Keeffe's New Mexico paintings. Not only could this type of work be useful to those wishing to learn more about the buildings depicted in her earlier works done in New York, this approach could also be used for her letters and other historical materials held by the Georgia O'Keeffe Museum. Making this context more easily available would be useful not only to researchers, but to visitors who wish to understand O'Keeffe's work within the context of her life.

Museums hold an important place in our society as one of the few places where experts and laypeople alike can go to learn about a topic of interest. Although the semantic web has yet to live up to Berners-Lee's vision of the World Wide Web of the future [1], it is already a powerful tool and geospatial information is uniquely suited to it.

#### References

- [1] Berners-Lee, T., Hendler, J. & Lassila, O. The semantic web. *Scientific american* **284**, 34–43 (2001).
- [2] Shadbolt, N., Berners-Lee, T. & Hall, W. The semantic web revisited. *IEEE intelligent systems* **21**, 96–101 (2006).
- [3] Hyvönen, E. et al. Museumfinland—finnish museums on the semantic web. Journal of Web Semantics 3, 224–241 (2005).
- [4] Parry, R., Poole, N. & Pratty, J. Semantic dissonances: Do we need (and to we understand) the semantic web? *Museums in a Digital Age* 96–106 (2008).
- [5] Georgia O'Keeffe Museum (2021). URL https://www.okeeffemuseum.org/.

- [6] GeoNames (2021). URL https://www.geonames.org/.
- [7] US Geological Survey (2021). URL https://www.usgs.gov/core-science-systems/ngp/cegis/linked-geospatial-data-system-architecture.
- [8] Battle, R. & Kolas, D. Geosparql: Enabling a geospatial semantic web. Semantic Web Journal 3, 355–370 (2011).
- [9] The Apache Software Foundation (2021). URL https://marmotta.apache.org/.
- [10] SemWebCentral (2021). URL https://github.com/SemWebCentral/parliament.
- [11] The Apache Software Foundation (2021). URL https://jena.apache.org/index.html.
- [12] W3C Semantic Web Interest Group (2021). URL https://www.w3.org/2003/01/geo/.
- [13] GeoNames (2021). URL https://www.geonames.org/6295630/earth.html.
- [14] Open Geospatial Consortium (2021). URL https://www.ogc.org/standards/geosparql.
- [15] Powell, L. J. & Varanka, D. E. A linked geodata map for enabling information access. Tech. Rep., US Geological Survey (2018).
- [16] Open Geospatial Consortium (2021). URL https://github.com/opengeospatial/ogc-geosparql/blob/maste
- [17] FransKnibbe (2020). URL https://github.com/opengeospatial/ogc-geosparql/issues/60.