Geospatial Reasoning with Shapefiles for Supporting Policy Decisions

Henrique Santos, James P. McCusker, and Deborah L. McGuinness

Rensselaer Polytechnic Institute, Troy NY, USA 12180

Abstract. Policies are authoritative assets that are present in multiple domains to support decision-making. They describe what actions are allowed or recommended when domain entities and their attributes satisfy certain criteria. It is common to find policies that contain geographical rules, including distance and containment relationships among named locations. These locations' polygons can often be found encoded in geospatial datasets. We present an approach to transform data from geospatial datasets into Linked Data using the OWL, PROV-O, and GeoSPARQL standards, and to leverage this representation to support automated ontology-based policy decisions. We applied our approach to location-sensitive radio spectrum policies to identify relationships between radio transmitters coordinates and policy-regulated regions in Census.gov datasets. Using a policy evaluation pipeline that mixes OWL reasoning and GeoSPARQL, our approach implements the relevant geospatial relationships, according to a set of requirements elicited by radio spectrum domain experts.

1   Introduction

Policies are commonly defined as decision-making assets that express one or more actions allowed or recommended under certain conditions. In the radio communications domain, policies are created to help manage the use of a limited electromagnetic spectrum. Many policies are location-specific, meaning that they are only applicable when the usage of the radio spectrum is to occur in specific geographic locations, as dictated by the policy. In the United States, the *National Telecommunications and Information Administration Manual of Regulations and Procedures for Federal Radio Frequency Management*[1] (*NTIA Redbook*) is a compilation of regulatory policies that define the conditions organizations, systems, and devices must satisfy to compatibly share radio spectrum while minimizing interference. Because policies in the NTIA Redbook regulate both commercial and federal spectrum usage, it is common to find military facilities, as well as regulations covering domestic and international locations.

The US Census Bureau publishes geospatial datasets about the United States, its territories, and points of interest, in its Census.gov data portal. The datasets contain high-definition polygons, usually in the `shapefile` [4] format, of many

---

[1] `http://bit.ly/NTIA_Redbook`

locations referred by radio spectrum policies. Although this format is a popular choice for encoding geographical entities, its main use-case is to support data interchange among geographic information systems (GIS). The `shapefile` format usually requires the use of a GIS to allow operations over the data, including calculations and queries. Therefore, it is not very suitable for integrating with ontology-based applications.

We present an approach to allow ontology-based applications to leverage geospatial data in formats not easily accessible or referred from within ontology constructs, and to use this data to perform geospatial calculations. We implemented the approach to support automated radio spectrum policy decisions. This is accomplished by representing Census.gov relevant polygons in the GeoSPARQL [5] vocabulary, and by defining OWL [9] classes that encode the policies' location rules. A policy evaluation pipeline that mixes OWL reasoning and GeoSPARQL leverages this model to elicit spatial relationships, providing high-definition spatial calculations. This approach was evaluated to perform well in terms of coverage of geospatial requirements, as elicited by domain experts. It is implemented as part of the Dynamic Spectrum Access (DSA) Policy Framework [14], which was developed to serve as a machine-readable policy repository to support increased automation of policy evaluations.

2 Transforming Census.gov `shapefiles`

The majority of location-sensitive policies in the NTIA Redbook refer to these types of locations: military facilities, States, or Country. Because the policies are originally authored in natural language, and targeted at spectrum managers, they refer to locations by their names (e.g. "Fairbanks", "Camp Parks"), without a comprehensive definition of the boundaries of such regions. To support automation of policy decisions, it becomes crucial to encode and leverage the polygons for these relevant locations.

The Census Bureau is the United States agency that serves as the nation's leading provider of quality data about its people and economy. Yearly, the agency publishes updated and authoritative geospatial datasets to provide meaning and context to statistical data the bureau produces. Largely published in the `shapefile` format, the published data[2] does include State boundaries and military installations, conveniently supporting our policy use-case. The State dataset is composed of 56 polygons, representing the 50 U.S. States, District of Columbia, plus 5 U.S. territories. The military installation dataset has 859 polygons, describing information about airports, laboratories, training areas, etc. In addition to the polygons, the datasets contain some minimal metadata about the locations, including a unique ID, and a legal name.

We have applied the *Semantic Extract, Transform, and Load-r* (SETLr [12]) to these datasets. SETLr orchestrates ETL pipelines by the use of a script in Turtle format that defines data sources, extract and transform processes, and

---

[2] `https://www.census.gov/programs-surveys/geography/geographies/mapping-files.html`

destination formats. SETLr was executed in both geospatial datasets to extract shapes' information and transform them into geographical features using the PROV-O [3] (`prov:Location`) and GeoSPARQL (`geo:Feature`, `sf:Geometry`) ontologies. Because each phase of the ETL pipeline in SETLr is defined as an RDF resource, the complete provenance of how these geographical features came to be is maintained, thereby supporting the explanation of policy decisions in more complex scenarios where multiple locations sources are involved.

3 Geospatial Reasoning on Radio Spectrum Policies

Geospatial reasoning is a crucial capability when evaluating policies. Many policies, including those that regulate radio spectrum usage, are only applicable when their specified location rules are satisfied. These locations include named locations that can be mapped to features from geospatial datasets, and polygons defined directly in the policy's rules. Either way, location rules need to be correctly evaluated, taking into consideration which polygons the policy regulates, as well as coordinates that are subject to evaluation (e.g. where a radio transmission is to occur).

We designed the DSA Policy Framework [14] to serve as a machine-readable, radio spectrum policy repository that can be used to automatically process radio transmission requests. The framework utilizes the World Wide Web Consortium's (W3C) OWL2 and PROV-O [3], and the Open Geospatial Consortium's (OGC) GeoSPARQL v1.0 standards as a modeling foundation of radio spectrum policies and involved entities. Figure 1 shows the RDF model of a transmission request within the DSA Policy Framework. Transmission requests are defined as `prov:Activity`, with the associated requester as a `prov:Agent`. Attributes that further characterize the transmission are represented using either PROV-O (including the location attribute) or a domain ontology. Coordinates in which requesters are located are represented as *Well-Known Text* (WKT) [2] string, and expressed using the `geo:asWKT` predicate.
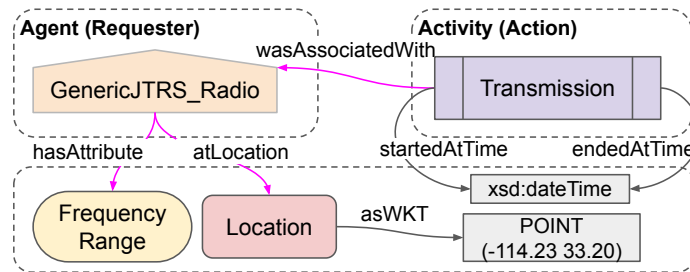

Fig. 1. The DSA request model

To allow the evaluation of the relationships between coordinates from transmissions and policy-regulated locations, we have pursued the representation of these locations as an OWL ontology where classes represent policy locations. To

exemplify this approach, we will use the second provision of the US91 policy from the NTIA Redbook, which reads (with adaptations):

> *"In the sub-band 1761-1780 MHz, Federal earth stations in the space operation service may transmit at the following 25 sites and non-Federal base stations must accept harmful interference caused by the operation of these earth stations: Fairbanks, Camp Parks, ..."*

Besides the policy text itself, which explicitly lists 25 sites where the policy is applicable, US91 is listed in the NTIA Redbook under the *United States table*, because it is only applicable in the US and not internationally. Listing 1 shows the representation of the involved locations for supporting this policy. Lines 1-7 define the `USLocation` class for expressing the entire United States land. This class is defined as a `prov:Location` and is a union of all States, District of Columbia, and territories from the appropriate Census.gov dataset, using the `geo:sfWithin` predicate from GeoSPARQL. Similarly, lines 9-16 extend this class to express the specific locations the above policy regulates, this time using features from the military facilities Census.gov dataset.

```
1 Class: USLocation
2   EquivalentTo:
3     prov:Location and (geo:sfWithin STATE_01 or
4                        geo:sfWithin STATE_02 or
5                        ...
6   SubClassOf:
7     prov:Location
8
9 Class: US91-2-c_Location
10   EquivalentTo:
11     USLocation and (
12     (geo:sfWithin value Fairbanks) or
13     (geo:sfWithin value CampParks) or
14     ...
15   SubClassOf:
16     USLocation
```

Listing 1. OWL expression of part of the US91 policy in Manchester syntax

3.1 Evaluating geospatial rules in policies

To evaluate coordinates in transmission requests with policy-regulated locations, we used the GeoSPARQL function predicates embedded in SPARQL queries, as seen in Listing 2. The implemented queries focus on the `within` and `distance` relationships. The queries infer triples in the format `:req_location geo:sfWithin :NAMED_LOCATION`, or as a distance attribute with the numerical distance as a value and in relation to some named location.

```
1 FILTER(geof:sfWithin({{WKT_STR}}^^geo:wktLiteral, ?wkt))
2 BIND(geof:distance({{WKT_STR}}^^geo:wktLiteral, ?wkt,
3   units:kilometer) AS ?distance)
```

Listing 2. GeoSPARQL statements to elicit select geospatial relationships.

The inferred triples are asserted back into the transmission request RDF model, which then gets reasoned over by an OWL reasoner. Using those inferred assertions, the location specified in the request can now be correctly reasoned to belong to one or more location classes, such as those in Listing 1. To exemplify this process, the coordinates for the request in Figure 1 are located in Arizona. Because US91's second provision does not include any Arizona locations, no triple linking the request location to one of the policy's locations would be inferred. But, a triple linking the request location to the State of Arizona would exist (`:req_location geo:sfWithin :STATE_04`). In this setting, the request location would be reasoned to belong to the `USLocation` class, but not to the `US91-2-c_Location` class, indicating that the transmission is to occur in the United States, but the second provision of US91 is not applicable.

Conversely, if the request in Figure 1 is modified to a coordinate within the "Fairbanks" named location, a triple `:req_location geo:sfWithin :Fairbanks` will exist. Therefore, the request location will be reasoned to belong to the `US91-2-c_Location` class, making the second provision of US91 applicable.

4   Evaluation

We worked with radio spectrum domain experts to elicit a set of geographical requirements that a machine-readable policy model needs to support. They appear in bold in the first column of Table 1. The table contains columns for Policy Representation and Request Evaluation. "Yes" indicates that the policy construct is either *Relevant* or it has been fully addressed and *Implemented*. "Partial" indicates that the current implementation meets a simplified requirement.

	Policy Representation		Request Evaluation	
Locations	**Relevant**	**Implemented**	**Relevant**	**Implemented**
Named locations	yes	yes	yes	yes
Relative locations	yes	partial	yes	partial
Polygons/Circles	yes	yes	yes	yes
Geographical rules				
Specific location	yes	yes	yes	yes
List of locations	yes	yes	yes	yes

Table 1. Geospatial semantics coverage

Most policies refer to locations by names or by coordinates (points, polygons, and circles), but sometimes a location is expressed in relation to another location. Currently, relative locations have been constrained to the ones expressed using the distance relationship. Geographical rules are defined in terms of the requester being in a location or a list of locations. Our approach implements these constructs using the `geo:sfWithin` predicate and OWL unions.

5   Related Work

The works in [11,13] proposed approaches for converting geospatial content to RDF, using mapping languages and ETL pipelines. The work in [6] allows the access of geospatial datasets, including `shapefiles`, using an ontology-based data access approach. Our conversion relied on SETLr, which enables the data conversion of geospatial data to RDF similar to the first two approaches, but also allows the maintenance of data transformation provenance. This maintenance is important in this use-case for supporting the explanation of policy decisions.

XACML 3.0, the eXtensible Access Control Markup Language [1], is a well-known policy language and *de facto* standard for representing attribute-based access control (ABAC) [10] policies and requests. Importantly, XACML provides a reference architecture for centralizing access control and a process model for evaluating requests against existing policies that inform the design of access control systems across domains and technologies. Thi [15] proposes an OWL-based extension to XACML to support a generalized, context-aware, role-based access control (RBAC) model, providing Spatio-temporal restrictions and conforming with the NIST RBAC standard [8]. Their work augments the XACML architecture with new functions and data types.

Our approach combines OWL, PROV-O, and GeoSPARQL to encode geospatial features, and an OWL reasoner to realize location class memberships. Our representation builds on previous work by matching the cross-domain policy expression semantics of XACML, extending it with the capacity to express rich Spatio-temporal restrictions, enabling the implementation of a wide variety of attribute-based policies across domains.

6   Conclusion

This paper presents an approach for leveraging geographical features, originally in `shapefiles`, to support policy decisions. In the radio spectrum domain, it is commonplace for policies to regulate the usage of the spectrum in specific locations, therefore requiring spatial reasoning to identify relationships between radio transmitters' coordinates and policy-regulated regions. This approach is an integral part of the DSA Policy Framework, which is functioning as a prototype policy management system in support of spectrum sharing operations.

Future work involves the research and development of the application of more spatial relationships, including relative locations. Besides, in other policy publications, we have encountered locations that are expressed in unusual shapes. These include paths, cones, and altitudes. More research is necessary to assess the impact in both modeling and reasoning, should we pursue this line of work. Finally, we are generalizing the approach to beyond radio spectrum policies by initially supporting practitioners from multiple domains in the creation of policies utilizing terminology and entities in domain knowledge graphs [7].

References

1. eXtensible Access Control Markup Language (XACML) Version 3.0. `http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html`
2. ISO/IEC 13249-3:2016 Information technology — Database languages — SQL multimedia and application packages — Part 3: Spatial
3. PROV-O: The PROV Ontology, `https://www.w3.org/TR/prov-o/`
4. ESRI Shapefile Technical Description (1998), `https://www.esri.com/Library/Whitepapers/Pdfs/Shapefile.pdf`
5. Battle, R., Kolas, D.: Enabling the geospatial Semantic Web with Parliament and GeoSPARQL. Semantic Web **3**(4), 355–370 (2012)
6. Bereta, K., Xiao, G., Koubarakis, M.: Ontop-spatial: Ontop of geospatial databases. Journal of Web Semantics **58**, 100514 (2019)
7. Falkow, M., Santos, H., McGuinness, D.L.: Towards a Domain-Agnostic Computable Policy Tool. In: ESWC 2021 Posters and Demos Track (2021)
8. Ferraiolo, D.F., Kuhn, D.R.: Role-Based Access Controls. In: 15th National Computer Security Conference (2009)
9. Grau, B.C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U.: OWL 2: The next step for OWL. Journal of Web Semantics **6**(4), 309–322 (2008)
10. Hu, V.C., Kuhn, D.R., Ferraiolo, D.F., Voas, J.: Attribute-based access control. Computer **48**(2), 85–88 (2015)
11. Kyzirakos, K., Savva, D., Vlachopoulos, I., Vasileiou, A., Karalis, N., Koubarakis, M., Manegold, S.: GeoTriples: Transforming geospatial data into RDF graphs using R2RML and RML mappings. Journal of Web Semantics **52-53**, 16–32 (2018)
12. McCusker, J.P., Chastain, K., Rashid, S., Norris, S., McGuinness, D.L.: SETLr: the semantic extract, transform, and load-r. Tech. Rep. e26476v1, PeerJ Inc. (2018)
13. Patroumpas, K., Alexakis, M., Giannopoulos, G., Athanasiou, S.: TripleGeo: an ETL Tool for Transforming Geospatial Data into RDF Triples. In: EDBT/ICDT Workshops. CEUR Workshop Proceedings, vol. 1133, pp. 275–278 (2014)
14. Santos, H., Mulvehill, A., Erickson, J.S., McCusker, J.P., Gordon, M., Xie, O., Stouffer, S., Capraro, G., Pidwerbetsky, A., Burgess, J., Berlinsky, A., Turck, K., Ashdown, J., McGuinness, D.L.: A Semantic Framework for Enabling Radio Spectrum Policy Management and Evaluation. In: The Semantic Web – ISWC 2020
15. Tran Thi, Q.N., Dang, T.K.: X-STROWL: A generalized extension of XACML for context-aware spatio-temporal RBAC model with OWL. In: Seventh International Conference on Digital Information Management (ICDIM). pp. 253–258 (2012)