Serving Ireland's Geospatial Information as Linked Data

Christophe Debruyne[1], Éamonn Clinton[2], Lorraine McNerney[2],
Atul Nautiyal[1] and Declan O'Sullivan[1]

The ADAPT Centre for Digital Content Technology, Trinity College Dublin, Dublin 2, Ireland
{debruync,nautiyaa,declan.osullivan}@scss.tcd.ie
Ordnance Survey Ireland, Phoenix Park, Dublin 8, Ireland
{eamonn.clinton,lorraine.mcnerney}@osi.ie

Abstract. We present `data.geohive.ie`, which aims to provide an authoritative platform for serving Ireland's national geospatial data, including Linked Data. Currently, the platform provides information on Irish administrative boundaries and was designed to support two use cases: serving boundary data of geographic features at various level of detail and capturing the evolution of administrative boundaries. We report on the decisions taken for modeling and serving the data such as the adoption of an appropriate URI strategy, the development of necessary ontologies, and the use of (named) graphs to support aforementioned use cases.

Keywords. Geospatial Data, Linked Data, Ontology Engineering

1   Introduction

In 2014, Ordnance Survey Ireland (OSi) delivered a newly developed spatial data storage model known as Prime2 [1]. With Prime2, OSi moved from a traditional map-centric model towards an object-oriented model from which various types of mapping and data services can be produced. Prime2 and the associated workflows furthermore designed governance practices to cope with evolution of spatial objects in their model. The system currently holds information of over 45,000,000 spatial objects (road segments, buildings, fences, etc.), of which some have more than one representation. These objects are stored in an Oracle Spatial and Graph database.[1]

At the same time, the OSi aims to adopt Linked Data [3] to enable third parties to explore and consume some of OSi's *authoritative* datasets.[2] We report on our current progress demonstrating the possibility of taking the Prime2 model as a basis for publishing OSi's geospatial data as Linked Data on the Web whilst adhering to best practices in the domain of geospatial information (e.g., by examining the work by the Ordnance Survey UK [4]). We start by publishing Ireland's boundary data, which

---

[1] `https://www.oracle.com/database/spatial/`

[2] Some of this data is also available via the Irish Government portal via `data.gov.ie`, but is currently not available as Linked Data. The goal is furthermore to investigate the feasibility of having a portal refer to OSi's data rather than hosting (and therefore also duplicating) it.

have been made available by OSi's Open Data release[3], taking into account two use cases: i) providing different "generalizations" (i.e. different levels of detail) of the boundaries and ii) capturing the evolution of boundaries, e.g. as ordered by Statutory Instruments. The main contributions of this paper are the decisions made for capturing and representing aforementioned information in RDF.

2 Approach and Implementation

With Prime2, OSi adopted an object-oriented model for capturing geospatial data. In this model, a clear distinction is made between a geographical object (identified by a GUID) and their representations. The distinction between objects and their representations is argued to be important [2], but in literature the terms *geographic features* and *geometries* are used. The geometry of a feature can evolve over time, and these changes do not have an impact on the feature. In other words, the geometry of a feature is "merely" an attribute. Prime2 will drive some of the design decisions we made for the development of the Linked Data platform.

2.1 URI Strategy

Unlike datasets that have been created at a specific time and for a specific purpose, such as CENSUS data, OSi's geospatial information is not static in nature; it does not make sense to include variables such as a creation date in the URIs of resources. Since each object is assigned a GUID, these can be used to create *opaque* URIs. We, however, have decided to include the type of geographical feature in the URI as to provide developers and consumers some idea of the nature of the entity the URI is referring to. This will not pose a problem as Prime2 *prescribes* that features that change of type (e.g. a convent becoming a hospital) are considered as new features and are therefore assigned new GUIDs.

2.2 Describing the Features and their Geometries

Since we have not found suitable ontologies for appropriately annotating the different boundaries (11 types in total; Baronies, Counties, County Councils, etc.), we decided to create a new ontology[4] that extends GeoSPARQL[5]. Ryan et al. noted some differences between concepts related to Ireland's geographic features and Linked Data datasets such as DBpedia and GeoNames [5]. Other problems include distinct definitions for "town lands" and "counties" and the absence of an ontology for describing "county council", amongst others. GeoSPARQL is both an ontology for describing geographical features and their geometries and defines predicates for spatial queries in SPARQL, making it a suitable candidate for our platform. Subclasses of the concept `geo:Feature` were introduced for each type of administrative boundary we serve.

GeoSPARQL supports the distinction between features and geometries. Since a

---

geometry is an attribute of a feature in the same way a name is an attribute of a person, we have, for the time being, chosen not to provide a URI for geometries. The geometries of a feature have thus to be accessed via a feature with `geo:hasGeometry`. Geometries are available in three levels of detail: generalized up to 100, 50 and 20 meters, which are stored in different (named) graphs. The default graph contains the features, labels in English and Gaelic (whenever available), and their representations generalized up to 100 meters (and are thus smaller in bandwidth). The generalizations up to 50 and 20 meters each have their own named graph.

Finally, Prime2 captures the geometries using the Irish Transverse Mercator (ITM) coordinate system. At an international level, however, World Geodetic System 84 (or WSG 84) is the standard used in cartography and navigation (amongst others). As OSi wishes to encourage the uptake of WGS 84 within Ireland, a decision was made to serve the geometries in WSG 84 only; third parties can themselves rely on services to transform the data between coordinate systems.

2.3 Capturing the Evolution of Geometries

Next we aim to capture the evolution of boundaries. Though rare, administrative boundaries can change with so called Statutory Instruments.[6] Statutory Instruments are available on the Web and are accessible via a URI, making it possible to relate the evolution of boundaries with these instruments. To capture the evolution of boundaries, we have chosen to extend PROV-O[7] with a new `prov:Activity` called "Boundary Change", which is informed by a new `prov:Entity` called "Statutory Instrument".[8] Prior versions of features and their geometries are captured in a separate graph. Note that geometries do not have a URI, but can be discovered via the feature.

3 The Platform

Objects are stored in an Oracle Spatial and Graph instance according to the Prime2 data model. RDF graphs are created by means of several R2RML[9] mappings that relate tables of the database with predicates in aforementioned ontologies. Those triples, currently 831,562 in total, are then loaded into a triplestore that supports Geo-SPARQL. To avoid an excessive load on the server, we have currently chosen to limit access to the SPARQL endpoint and set up a Triple Pattern Fragments (TPF) server [6] instead. A TPF server basically returns a result set for simple triple patterns and it is up to a TPF client to compute the result of a SPARQL query. A limitation is that TPF does not (yet) support the geospatial predicates provided by GeoSPARQL and users therefore have no way to exploit these on the platform. The platform furthermore hosts the boundary datasets as dumps and hosts the two aforementioned ontologies for Irish administrative boundaries according to Linked Data principles.

---

[6] An example of a Statutory Instrument altering borders between counties can be found here: `http://www.irishstatutebook.ie/eli/1994/si/114/made/en/print#`

[7] `https://www.w3.org/TR/prov-o/`

[8] `http://ontologies.geohive.ie/osiprov`

[9] `https://www.w3.org/TR/r2rml/`

4   Conclusions and Future Work

We presented our ongoing work in creating `data.geohive.ie` for serving Ireland's geospatial information as an authoritative Linked Data dataset on the Web. We currently serve 11 types of administrative boundaries and focused on two use cases: serving different levels of detail and capturing the evolution of boundaries.

One of the main limitations of this platform is that it does not yet serve meaningful data for the latter. This is due to the fact that the boundary dataset is quite static and prior versions of administrative boundaries have not (yet) been entered in the Prime2 data model. Though simulated, we should have access to such data to validate that aspect of our approach. Future work described below would provide datasets to validate our approach to capturing the evolution of geometries. Another limitation is that users are currently unable to use the spatial predicates via Triple Pattern Fragments, as the SPARQL endpoint is not (yet) made available to the public. The OSi will first monitor the use of the platform before deciding whether this can be made available.

We aim to incorporate other administrative boundaries and other types of features in the future. For the first, we are looking at small areas that have been used for the 2011 CENSUS, allowing us to create links with CENSUS 2011 information published at `data.cso.ie`. For the latter, we will look into the inclusion of features that are *not open* and are made available via commercial licenses to certain parties, such as the geometries of buildings. Serving these datasets as Linked (Closed) Data would thus require the investigation of access control mechanisms. Approved construction works may change the geometry of buildings and are captured in the Prime2 model. This dataset could thus be used to validate our approach to capturing geometry evolution.

References

1. Prime2: Data Concepts and Data Model Overview. Tech. rep., Ordnance Survey Ireland (2014), `http://www.osi.ie/wp-content/uploads/2015/04/Prime2-V-2.pdf`
2. Battle, R., Kolas, D.: Enabling the geospatial semantic web with parliament and Geo-SPARQL. Semantic Web 3(4), 355–370 (2012)
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. Int. J. Semantic Web Inf. Syst. 5(3), 1–22 (2009)
4. Goodwin, J., Dolbear, C., Hart, G.: Geographical linked data: The administrative geography of Great Britain on the semantic web. Transactions in GIS 12, 19–30 (2008)
5. Ryan, C., Grant, R., Carragáin, E.Ó., Collins, S., Decker, S., Lopes, N.: Linked data authority records for Irish place names. Int. J. on Digital Libraries 15(2-4), 73–85 (2015)
6. Verborgh, R., Vander Sande, M., Hartig, O., Van Herwegen, J., De Vocht, L., De Meester, B., Haesendonck, G., Colpaert, P.: Triple pattern fragments: A low- cost knowledge graph interface for the Web. J. Web Sem. 37, 184–206 (2016)