Publishing and Consuming Irish Administrative Boundaries as Linked Data

Christophe Debruyne(✉), Atul Nautiyal, and Declan O'Sullivan

ADAPT Centre for Digital Content Technology,
Trinity College Dublin, Dublin 2, Ireland
{debruync,nautiyaa,declan.osullivan}@scss.tcd.ie

Abstract. We report on the Linked Data platform developed for the administrative boundaries governed by the Ordnance Survey Ireland (OSi), as they wished to serve this data as an authoritative Linked Open Data dataset on the Web. To implement this platform, we have adopted best practices and guidelines from the industry and academia. We demonstrate how this dataset can be combined with other datasets to add a spatial component to information. We believe that the publication of this dataset not only provides opportunities for third parties (including scholars) in their activities, but that this outcome of this initiative is of importance, as the OSi made the authoritative dataset available. With the current platform deployed, future work will include the inclusion of other (closed) datasets and the investigation of access mechanisms.

Keywords: Linked data · Geospatial information · GeoSPARQL

1 Introduction

Linked Data [2] refers to both an initiative and a set of best practices and guidelines to publish and interlink data on the Web using standardized Web technologies such as HTTP URIs, RDF and SPARQL. Important is the availability of authoritative datasets published as Linked Data that allows one to interlink information, create novel applications or support third parties in their activities such as scholars analyzing datasets. An example of the inclusion of an authoritative dataset as RDF into the Linked Data Web is Linked Logainm [4], where a set of Irish place names were related with their geographic counterpart in GeoNames[1] and DBpedia[2] using the SILK Link Discovery Network [3]. The Ordnance Survey Ireland, Ireland's National Mapping Agency, embarked on an initiative to serve an authoritative boundaries dataset they govern as Linked Data. In this paper, we elaborate on OSi's Linked Data platform and demonstrate how this dataset can be used with other datasets for scholarly activities.

[1] http://geonames.org/.
[2] http://dbpedia.org/.

B. Bozic et al. (Eds.): CHDDH 2016, IFIP AICT 482, pp. 113–117, 2016.
DOI: 10.1007/978-3-319-46224-0_11

2 OSi's Linked Data Platform

The platform is available at http://data.geohive.ie. An important distinction has to be made between *geographic features* and their *geometries* [1]. The first denotes things such as building, counties, forests, and the latter their geometric representation. For the first, we have developed an ontology[3] for the administrative boundaries that have been made available as open data through Ireland's New National Mapping Agreement[4]. Features such as Barony and County were introduced as subclasses of GeoSPARQL[5]'s concept of Feature.

Since we argue that a geometry is "merely" 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 geometries with a URI. 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 m. The level of detail has an impact on bandwidth and rendering, amongst others. An example of how descriptions of features are presented in HTML is shown in Fig. 1.

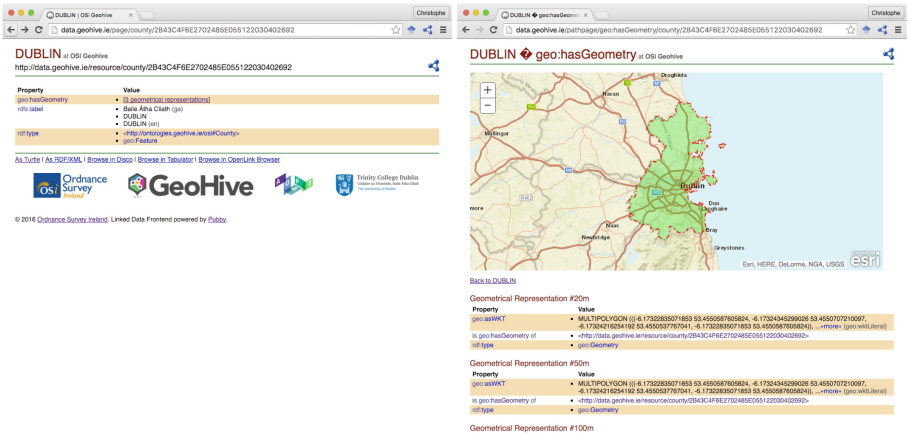

Fig. 1. Description in HTML of County Dublin on the left and its three geometries – with the one generalized up to 100 m drawn on a map – on the right.

We have also decided to separate *non-information resources* from *information resources*, the first being things and the latter being documents describing these things, by giving them different HTTP URIs. For instance, the County Dublin is identified with the URI x, described by the HTML document with URI y and described by an RDF document with URI z. Obtaining the representation that one needs is done with a technique called *content negotiation*.

[3] http://ontologies.geohive.ie/osi.
[4] http://www.osi.ie/news/new-mapping-agreement/, last accessed April 5, 2016.
[5] http://www.opengis.net/ont/geosparql.

To avoid an excessive load on the server, we have chosen to limit access to the SPARQL endpoint and set up a Triple Pattern Fragments (TPF) server [5] 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. The platform furthermore hosts the boundary datasets as dumps and hosts simple ontologies for Irish administrative boundaries according to Linked Data principles.

3 Consuming Ireland's Boundary Data

The administrative boundaries that are currently available as Linked Data are: City and County Council, City Council, Civil Parish, County Council, Electoral Division, Local Electoral Area, Municipal District, Rural Area, Barony, County, and Townland. We note that City, County, and City and County Councils are indeed three separate entities.

To demonstrate how the boundary data can be used, we will combine it with the 2011 Census data.[6]. We will look at the number of people in private households by size in "CTY areas in Ireland"[7]. This concept corresponds with the union of City, County, and City and County Councils in the OSi dataset. There are 34 CTYs in the census data. The OSi data has 26 County Councils, 3 City Councils and 2 City and County Councils. These numbers seem not to add up, but it is important to note that the data was collected in 2011 and the counties of Tipperary North and Tipperary South were merged into County Tipperary in 2014. The census has also split the city and county of the 2 City and County Councils considered as administrative boundaries by the OSi.

The CSO dataset contains observations for each area. One type of observation collected is the number of people living in households of different sizes. By retrieving those with the query below and asserting `owl:sameAs` statements between the correspondences, one can formulate, for instance a query to retrieve the total numbers of people living households of 8 people or more. These can then be plotted on a map using OSi's boundary data, as shown in Fig. 2.

```
prefix qb: <http://purl.org/linked-data/cube#>
DESCRIBE ?x WHERE {
?x a qb:Observation .
?x qb:dataSet <http://data.cso.ie/census-2011/dataset/persons-in-houses-by-size/cty>.
}
```

This demonstrates that OSi's authoritative boundary data can be easily combined with other datasets and add a spatial component for scholars to explore. While not demonstrated in this paper, the geospatial infrastructure allows one also to retrieve information via the geometries (e.g., "retrieve all civil parishes in this square").

[6] Available as Linked Data on http://data.cso.ie/.
[7] See http://data.cso.ie/census-2011/page/classification/areas/CTY.

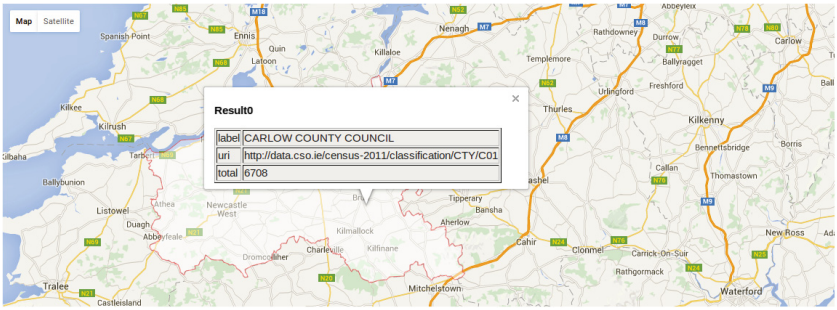

Fig. 2. Plotting the results of the query on a map.

4 Conclusions and Future Work

In this paper, we reported on the development of a Linked Data Platform for Ireland's Administrative Boundaries for and provided by the Ordnance Survey Ireland, who are the custodians of that data. As they are the custodians, the dataset that has been published is regarded as authoritative. We have demonstrated how this data can be combined with other datasets to This demonstrates that OSi's authoritative boundary data can be easily combined with other datasets, which can facilitate data exploration for, amongst others, scholars.

Current limitations are the absence of "versions" of administrative boundaries and the limited availability to the SPARQL endpoint. Data about boundary evolution, though addressed from a conceptual point of view and simulated, cannot be served as they are not (yet) stored in OSi's technology stack. TPFs do not provide support for all SPARQL queries and GeoSPARQL's spatial predicates. Access mechanisms to the SPARQL endpoint will be investigated.

Acknowledgments. The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund. We thank the Ordnance Survey Ireland (OSi) for permitting us to use their boundaries dataset for the purposes of this study. Within OSi, we are especially grateful for the input and domain expertise provided by Lorraine McNerney and Éamonn Clinton.

References

1. Battle, R., Kolas, D.: Enabling the geospatial semantic web with parliament and geosparql. Seman. Web **3**(4), 355–370 (2012)
2. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. Int. J. Seman. Web Inf. Syst. **5**(3), 1–22 (2009)
3. Isele, R., Jentzsch, A., Bizer, C.: Silk server - adding missing links while consuming linked data. In: Hartig, O., Harth, A., Sequeda, J. (eds.) Proceedings of the First International Workshop on Consuming Linked Data, Shanghai, China, November 8, 2010. CEUR Workshop Proceedings, vol. 665. CEUR-WS.org (2010)

4. Ryan, C., Grant, R., Carragáin, E.Ó., Collins, S., Decker, S., Lopes, N.: Linked data authority records for Irish place names. Int. J. Digit. Libr. **15**(2–4), 73–85 (2015)
5. Verborgh, R., Vander Sande, M., Hartig, O., Herwegen, J., Vocht, L., 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)