Christophe Debruyne, Alan Meehan, Éamonn Clinton, Lorraine McNerney, Atul Nautiyal, Peter Lavin and Declan O'Sullivan. Ireland's Authoritative Geospatial Linked Data. In Claudia d'Amato, Miriam Fernández, Valentina A. M. Tamma, Freddy Lécué, Philippe Cudré-Mauroux, Juan F. Sequeda, Christoph Lange, and Jeff Heflin, editors, The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part II, volume 10588 of Lecture Notes in Computer Science, pages 66–74. Springer, 2017

Ireland's Authoritative Geospatial Linked Data

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

[1]ADAPT Centre, Trinity College Dublin, Dublin 2, Ireland
{first.last}@adaptcentre.ie
[2]Ordnance Survey Ireland, Phoenix Park, Dublin 8, Ireland
{first.last}@osi.ie

Abstract. Data.geohive.ie aims to provide an authoritative service for serving Ireland's national geospatial data as Linked Data. The service currently provides information on Irish administrative boundaries and the boundaries used for the Irish 2011 census. The service is designed to support two use cases: serving boundary data of geographic features at various level of detail and capturing the evolution of administrative boundaries. In this paper, we report on the development of the service and elaborate on some of the informed decisions concerned with the URI strategy and use of named graphs for the support of aforementioned use cases – relating those with similar initiatives. While clear insights on how the data is being used are still being gathered, we provide examples of how and where this geospatial Linked Data dataset is used.

Keywords. Geospatial Linked Data, Ordnance Survey Ireland
Resource type: Dataset
Permanent URL: http://purl.org/geohive

1 Introduction

The Ordnance Survey Ireland (OSi), Ireland's national mapping agency, aims to adopt Linked Data to enable third parties to explore and consume some of OSi's authoritative datasets. In [5], we reported on how the OSi's object-centric relational database, called Prime2 [1], was used to publish administrative boundary datasets according to best practices and guidelines for geospatial Linked Data. The service was developed to support two use cases: i) providing the boundary detail in varying levels of detail and ii) capturing the evolution of boundaries. In this paper, we provide more details on the dataset [4], and its value and potential impact in the context of Ireland.

2 Related Work

Shadbolt et al. highlighted the importance of location in data and its role in interlinking and aligning datasets [18]. This is certainly the case for government data, which often reports numbers that are related to certain territories (administrative units, juris-

dictions, etc.). The Linked Data Web has numerous geographic datasets; GeoNames and LinkedGeoData[1] (which cover a vast part of the world) and Ordnance Survey Linked Data[2] (for the UK), just to name a few. Except for the latter, many of these geographic datasets are not authoritative in nature, nor are they necessarily accurate. LinkedGeoData, for example, uses the information collected by the OpenStreetMap[3] project, which itself is an open environment in which volunteers collaboratively create a geospatial knowledge base. Though OpenStreetMap is quite accurate compared to official sources [9], its coverage has been shown to be incomplete [14]. Though the data provided by LinkedGeoData might be good for a lot of applications; one may wish or need to avail of authoritative datasets with legal weight. One can thus see the potential and added value of publishing and linking with *authoritative* geospatial data.

The Ordnance Survey of Great Britain was one of the first to publish some of their geospatial data on the Web [8]. While this is a great example of publishing authoritative geospatial Linked Data, in our opinion, it is unfortunate that they have not adopted a standard for representing features, spatial relations, and representation of geometries. Instead, they rely on a bespoke ontology. Reasoning over their geospatial data either requires relying on rules for that bespoke schema, or mapping the data onto standardized vocabularies such as OGC GeoSPARQL [17], for which implementations exists.

Other countries are looking at publishing their authoritative geospatial information on the Web as well. One such example is the Cadaster[4] in The Netherlands, which is driven by the public administration. [2] proposed vocabularies and an approach for serving geographic reference data for the French national mapping agency. In the EU, the INSPIRE directive (Infrastructure for Spatial Information in Europe) aims to standardize Spatial Data Infrastructures across Europe. In order for one to discover, access and visualize geospatial information in a homogenous manner across Europe, the directive prescribes metadata formats, services, etc. that each member state has to comply with. [16] proposed to map INSPIRE onto GeoSPARQL to provide an RDF perspective on such data and applied their method in the context of Greece.

3 Approach

In this section we elaborate on how the OSi's geospatial information has been organized and how this has been delivered to agents.

3.1 URI Strategy

Coming up with adequate URI strategies for publishing 5-star Linked Open Government Data on the Web is challenging, especially when one has to take into account the difference in governance practices, heterogeneity, etc. across different government bodies. A URI strategy for geospatial data has been proposed in [19], which was

[1] `http://www.geonames.org/` and `http://linkedgeodata.org/`
[2] `http://data.ordnancesurvey.co.uk/`
[3] `https://www.openstreetmap.org/`
[4] `http://www.pilod.nl/`, and `http://almere.pilod.nl/sparql`

based on a more generic URI strategy for The Netherlands [15].

In the case of the OSi, the term "dataset" in "boundaries dataset" is actually a misnomer when referring to administrative boundaries in Ireland. This particular dataset is a dynamic dataset that evolves over time, unlike datasets that are created at a particular point in time such as census data. While progress has been made since the start of this project on drafting a URI strategy for the Irish Government's open data initiative [11,12], early discussions encouraged the inclusion of attributes such as creation date in the HTTP URIs. This approach would not have suited the OSi as this necessitated the creation of datasets for each change. This in turn would have complicated the governance of links between these datasets, and also the governance of links to the OSi datasets by 3rd parties. In conjunction with the Department of Public Expenditure and Reform (DPER) and the Central Statistics Office (CSO), we have decided to use a subset of the recommended attributes, allowing us to still be in line with most of the recommendations that were then put forward.

Currently, URIs, for the resources that the OSi are the custodians of, follow the following pattern: http://data.geohive.ie/{**type**}/{**concept**}/{**GUID**}, where:

- The *domain* follows the two recommendations formulated by [15]: solely be used for the publication of OSi's geospatial information and not include the name of any organization, as they may evolve over time.[5]
- *Type* can take any of the following values: "resource" for the HTTP URI of a resource, and "page" and "data" for that resource's HTML and RDF documents respectively.
- *Concept* and *GUID*: with Prime2, all features are assigned a *GUID*. Therefore, although we would have been able to create fully opaque URIs by only providing the GUID, we have chosen to provide a hint of what this resource is about by providing a label referring to that resource's class in *concept*.

Concerning the GUIDs, we note that Prime2 provides governance rules that prescribe how features may evolve over time. One of these rules prescribes that features do not change in nature. When a hospital is transformed into an apartment building, for example, it is considered a new feature (and therefore has a different GUID) that happens to have the same geometric representation.

Finally, one important decision that we have made concerning our URI strategy was not to provide URIs for the geometries. A clear distinction is made between a geographical feature (such as a county), and its geometry (such as its boundary represented by a polygon). When adopting ontologies such as GeoSPARQL (see the next section), two distinct classes reflect this distinction. This means that instances of these classes can be identified with a URI. In practice, we notice that users abuse the boundaries and use them as the identifier of the feature. In other words, they would refer to the county's boundary as the county, rather than referring to the resource rep-

[5] GeoHive is an initiative by OSi to provide easy access to publicly available authoritative spatial data. The same top-level domain was used for the publication of their Linked Data. We have not chosen to adopt a sub-domain under osi.ie as by 2017, several government bodies including the OSi will be merged and change (domain) names. GeoHive will remain.

resenting the county. To avoid this problem for OSi's Linked Data, we have decided not to provide URIs to geometries and publish them as blank nodes.

3.2 Knowledge Organization: Different Representations

The distinction between a geographic feature and its geometry (or even geometries) is argued to be important [3]. The geometry of a feature can evolve over time – e.g., due to coastal erosion, and these changes do not have an impact on the feature. In other words, the geometry of a feature is "merely" an attribute.

Since we have not found suitable ontologies for appropriately annotating the different administrative boundaries (e.g., Counties and Electoral Divisions) in an Irish context, we decided to create a new ontology[6] that *extends* GeoSPARQL.[7] GeoSPARQL is an ontology for describing geographical features and their geometries. It also defines predicates for spatial queries in SPARQL, making it a suitable candidate for our service. Subclasses of the concept `geo:Feature` were introduced for each type of administrative boundary we serve.

Finally, OSi's bespoke information system captures the geometries using the Irish Transverse Mercator (ITM) coordinate system. At an international level, however, World Geodetic System 84 (or WGS 84) is the standard used in cartography and navigation. As OSi also wishes to encourage the uptake of WGS 84 within Ireland, a decision was made to serve the geometries in WGS 84 only; third parties can themselves rely on services to transform the data between coordinate systems. We use the Well-known Text (WKT) markup language for representing the geometries.

Our first use case was to provide boundary data with different levels of detail (or "resolutions"). The polygons are generalized up to 20, 50 and 100 meters. Higher resolutions provide more detail but require more data transfer. Different resolutions are used for different purposes; the Irish census uses 20m resolutions and 100m resolutions for information exchange at a European level, for instance

We generate instances of `geo:Geometry` for each resolution and store them in dedicated graphs (one for each resolution). The feature and its resolutions are related with `geo:hasGeometry`. A `geo:defaultGeometry` predicate is also declared between the feature and its 20m boundary data, as per best practice. Moreover, if two features happen to have geometries which are identical polygons, we do not reuse that geometry. Instead, we create two geometries that happen to have the same polygon (WKT literal). We then attach provenance information to each of these geometries. This is necessary as each feature (and its geometry) may have a different change history (see Section 3.3).. Finally, links from features to resources in external Linked Data datasets are stored in a separate named graph.

3.3 Knowledge Organization: Evolution of Geometries

Our second use case was to support capturing the evolution of boundaries. Though

[6] `http://ontologies.geohive.ie/osi#`

[7] We note that we did not consider reusing vocabularies that were not built on standards (e.g., [8]) or that were developed for another context (e.g., see [2] for France).

they are rare for administrative boundaries, they are ordered by so called Statutory Instruments. 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 [13] 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 separate graphs.

At the present time, OSi's database only contains current versions of administrative boundary data and does not contain any historical record of versions that may have existed in the past (i.e., prior to the release of Prime2 in 2014). OSi's database has not yet started ingesting prior (versions of) administrative boundary data before its release in 2014. We therefore have to rely on simulations, using geometries related to buildings, to demonstrate the feasibility of this approach. Geometries that are related to buildings have a much higher churn, but are not part of OSi's open data.

One can argue that capturing all provenance information related to boundary changes into one graph (per resolution) results in – over time – very large graphs. Indeed, another approach would have been to capture each change in dedicated graphs, which is the approach adopted by the Dutch public administration (see Section 4). The latter, however, would require the formulation of queries over different named graphs. Our approach was informed by the fact that use cases for retrieving the history of geometries are specific (e.g., of interest to building planners), which makes us believe that simpler queries will be favored at the expense of query execution time.

4 Discussion

In this section, we discuss the evaluation criteria as outlined by the ISWC 2017 call for resources track papers [10].

On Potential Impact. The resource is sufficiently general to be applied in many domains and scenarios, and this supports the arguments which will be made about its reusability (see "On Reusability"). The resource provides an authoritative source for use when adding a geospatial dimension to other datasets. The resource can be used by, inter alia, other Linked Data initiatives that are ongoing or emerging in government entities across Ireland. Therefore, its impact is more societal in nature.

The design and approach used in developing the resource has been compared to the state of the art. It has also been presented (at a seminar,) to representatives of other public administrations who have started similar initiatives (e.g. The Netherlands and Flanders, Belgium).[9] Ireland and The Netherlands have adopted different approaches to organizing the history of features and geometries using PROV-O, and we hope, over time, to inform each other of insights gained.

On Reusability. Shadbolt et al. [18] have already provided the motivation for, and established the usefulness of geospatial data for aligning, exploring and analyzing

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

[9] http://www.pilod.nl/wiki/Linked_Data_Seminar_-_December_2,_2016

data in many domains and scenarios. Furthermore, Ireland's Department of Public Expenditure and Reform has funded two projects via their Open Data Engagement Fund. The first project was to inform the public on how to add an authoritative geospatial dimension to CSV files on their open data portal [7]. The second project organized seminars on publishing and interlinking Linked Data with the resource. Data.cso.ie – an initiative between the CSO and the Insight Centre for Data Analytics – is a Linked Data Service for the census 2011 (and soon 2016) results. We have sent to data.cso.ie a set of links between their boundary identifiers and our administrative units. It is hoped they will deploy those links at the same time as they publish the 2016 results. With regard to the 2016 census boundaries; the ontology is straightforward to extend and we will adopt a similar approach for generating Linked Data for those boundaries as soon as the census 2016 polygons have been approved for publication. We have anecdotal evidence that various groups are using the resource. As an example, the Chronic Disease Informatics Group (CDIG) in Trinity College Dublin is using the datasets to relate observations (weather, pollution, etc.) to particular administrative boundaries in an effort to identify triggers for particular diseases.

On Design and Technical Quality. In the previous section, we provided details on our URI strategy, adoptions and extensions of standardized vocabularies, as well as informed decisions on knowledge organization. All of these are informed by best practices in other public administrations and provide for both the evolution of geometries as well as multiple representations thereof. The reuse of those standardized vocabularies allow agents, both human and computer-based, to avail of those predicates with existing tools; especially using the spatial predicate provided by GeoSPARQL. We furthermore like to stress our informed decision not to provide HTTP URIs to the geometries, as they are "merely" attributes of a feature that can evolve over time and to encourage users to link to entities rather than their "shapes".

Metadata in VoID about the boundaries dataset has been generated for the whole dataset, but also for specific subsets (e.g., County Councils of Ireland) that can be found on the resource's website. The whole dataset and its VoID dataset description have been made available on DataHub.

Both the URIs of resources and our ontologies resolve to human and machine-readable representations via content negotiation. In addition, the HTML pages of the resources even plot the geometries on OSi's basemaps.

On Availability. The dataset is available on http://data.geohive.ie/ and on DataHub.io, both of which provide links to dumps. Data is provided under a Create Commons Attribution 4.0 International license (CC BY 4.0) which is documented in both the HTML and in the dataset description using VoID. URIs resolve to either HTML pages or RDF serialization by means of content negotiation. The OSi has decided not to provide access to a GeoSPARQL endpoint, but instead refers to a Triple Pattern Fragments (TPF) [20] Server and Client are provided. We also provide a TPF client that has been extended with GeoSPARQL functions to allow users to query over the geometries [6]. The resource has furthermore been published on DataHub (with appropriate license information) and uses and extends standardized vocabularies such as GeoSPARQL and PROV-O. This enhances its reusability in other contexts.

Organizations are subject to changes and this impacts on web domain names used; OSi is no exception. At the end of 2017, OSi will merge with the Property Registration Authority of Ireland (PRA) and the Valuation Office (VO) to create Tailte Éireann, a new government body. Following this, mapping services, Prime2 and GeoHive will be under the remit of Tailte Éireann. Such a merging of bodies validates the decision to dedicate the domain name `data.geohive.ie` to the resource, a name not tied to any of those bodies, facilitating the sustainability of the resource.

5 Conclusions and Future Work

In this paper, we presented the authoritative boundaries dataset that has been made available as Linked Open Data with a CC BY 4.0 license. The data and ontologies developed for this dataset extend standardized vocabularies such a PROV-O and GeoSPARQL, facilitating its interoperability. Future work consists of extending the dataset with the boundaries used for the 2016 census and other (administrative) boundaries not yet included in this dataset. We aim to gather further insights into our approach for capturing the evolution of boundaries in a provenance graph and compare those with similar initiatives elsewhere (e.g., in The Netherlands).

Acknowledgements. 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.

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. Atemezing, G.A., Abadie, N., Troncy, R., Bucher, B.: Publishing reference geodata on the web: Opportunities and challenges for IGN France. In: Kyzirakos, K., Grütter, R., Kolas, D., Perry, M., Compton, M., Janowicz, K., Taylor, K. (eds.) Joint Proceedings of the 6th International Workshop on the Foundations, Technologies and Applications of the Geospatial Web, TC 2014, and 7th International Workshop on Semantic Sensor Networks, SSN 2014, co-located with 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Trentino, Italy, October 20, 2014. CEUR Workshop Proceedings, vol. 1401, pp. 9-20. CEUR-WS.org (2014)
3. Battle, R., Kolas, D.: Enabling the geospatial semantic web with Parliament and GeoSPARQL. Semantic Web 3(4), 355–370 (2012), http://dx.doi.org/10.3233/SW-2012-0065
4. Debruyne, C., Meehan, A., Clinton, E., McNerney, L., Nautiyal, A., Lavin, P., O'Sullivan, D. (2017): GeoHive Administrative Boundaries Dataset. figshare. https://doi.org/10.6084/m9.figshare.5235961
5. Debruyne, C., Clinton, E., McNerney, L., Nautiyal, A., O'Sullivan, D.: Serving Ireland's geospatial information as linked data. In: Kawamura, T., Paulheim, H. (eds.) Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with 15th International Semantic Web Conference (ISWC 2016), Kobe, Japan, October 19, 2016. CEUR Workshop Proceedings, vol. 1690. CEUR-WS.org (2016)
6. Debruyne, C., Clinton, E., O'Sullivan, D.: Client-side processing of GeoSPARQL func-

tions with Triple Pattern Fragments. In: Auer, S., Berners-Lee, T., Bizer, C., Capadisli, S., Heath, T., Janowicz, K., Lehmann, J. (eds.) Workshop on Linked Data on the Web co-located with 26th International World Wide Web Conference (WWW 2017). CEUR Workshop Proceedings, CEUR-WS.org (2017)
7. Debruyne, C., McGlinn, K., McNerney, L., O'Sullivan, D.: A lightweight approach to explore, enrich and use data with a geospatial dimension with semantic web technologies. In: Proceedings of the Fourth International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data. pp. 1:1-1:6. GeoRich '17, ACM, New York, NY, USA (2017), http://doi.acm.org/10.1145/3080546.3080548
8. Goodwin, J., Dolbear, C., Hart, G.: Geographical linked data: The administrative geography of Great Britain on the semantic web. Transactions in GIS 12(s1), 19–30 (2008)
9. Haklay, M.: How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. Environment and planning B: Planning and design 37(4), 682–703 (2010)
10. Lecue F., and Tamma, V.: ISWC 2017 Resources Track: Author and Reviewer Instructions. (2017), http://dx.doi.org/10.6084/m9.figshare.4679152
11. Lee, D., Cyganiak, R., Decker, S.: Open Data Ireland: Best Practice Handbook. Tech. rep., Insight Centre for Data Analytics, NUI Galway (2014), http://per.gov.ie/wp-content/uploads/Best-Practice-Handbook.pdf
12. Lee, D., Cyganiak, R., Decker, S.: Open Data Ireland: Open Data Publication Handbook. Tech. rep., Insight Centre for Data Analytics, NUI Galway (2014), http://per.gov.ie/wp-content/uploads/Open-Data-Publication-Handbook.pdf
13. McGuinness, D., Lebo, T., Sahoo, S.: PROV-O: The PROV ontology. W3C Recommendation, W3C (Apr 2013), http://www.w3.org/TR/2013/REC-prov-o-20130430/
14. Mooney, P., Corcoran, P., Winstanley, A.C.: Towards quality metrics for OpenStreetMap. In: Agrawal, D., Zhang, P., Abbadi, A.E., Mokbel, M.F. (eds.) 18th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, ACM-GIS 2010, November 3-5, 2010, San Jose, CA, USA, Proceedings. pp. 514–517. ACM (2010), http://doi.acm.org/10.1145/1869790.1869875
15. Overbeek, H., van den Brink, L.: Towards a national URI-Strategy for Linked Data of the Dutch public sector. Tech. rep., Kennis- en Exploitatiecentrum Officiële Overheidspublicaties & Geonovum (2013), http://www.pilod.nl/w/images/a/aa/D1-2013-09-19_Towards_a_NL_URI_Strategy.pdf
16. Patroumpas, K., Georgomanolis, N., Stratiotis, T., Alexakis, M., Athanasiou, S.: Exposing INSPIRE on the Semantic Web. J. Web Sem. 35, 53–62 (2015), http://dx.doi.org/10.1016/j.websem.2015.09.003
17. Perry, M., Herring, J.: GeoSPARQL - A Geographic Query Language for RDF Data, OGC (2012), http://opengeospatial.org/standards/geosparql
18. Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., Hall, W., schraefel, m. c.: Linked open government data: Lessons from data.gov.uk. IEEE Intelligent Systems 27(3), 16–24 (2012), http://dx.doi.org/10.1109/MIS.2012.23
19. van den Brink, L., Janssen, P., Quak, W., and Stoter, J.: Linking spatial data: semi-automated conversion of geo-information models and GML data to RDF. International Journal of Spatial Data Infrastructures Research 9, 59–85 (2014)
20. 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. 37-38, 184-206 (2016), https://doi.org/10.1016/j.websem.2016.03.003