Multiple Representation for Geospatial Linked Data

Weiming Huang
GIS Centre, Department of Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12, 223 62
Lund, Sweden
weiming.huang@nateko.lu.se

Eiður Eiðsson
Department of Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12, 223 62
Lund, Sweden
eidur.eidsson.715@student.lu.se

Lars Harrie
GIS Centre, Department of Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12, 223 62
Lund, Sweden
lars.harrie@nateko.lu.se

Abstract

Current techniques for visualising geospatial Linked Data are limited in terms of multiple representation, which has been studied for decades by cartographers and considered as a prerequisite for deriving appropriate geovisualisation applications. In order to alleviate this issue, this paper presents a work in progress, in which the multiple representation geospatial data are released as Linked Data, and linked to the Linked Data gazetteer GeoNames. In the study, we use extended INSPIRE draft RDF vocabularies to explicitly link multiple representations and their visualisation scales. The results show that the released multiple representation geospatial Linked Data can effectively enrich the geometric information for the test data in GeoNames, and provide better visualisation performance.

Keywords: geovisualisation, Linked Data, multiple representation, INSPIRE RDF vocabularies, scale

1 7Introduction

The amount of geospatial data published as Linked Open Data (LOD) has considerably increased in the last years. These geospatial LOD include both authoritative data (e.g. the Ordnance Survey Linked Data [1]) and crowd-sourced data (e.g. LinkedGeoData[2]). Geospatial information is widespread and can serve as nexuses that interconnect data from different sources (Goodwin *et al*. 2008), thus the geospatial datasets reside in central places in the LOD cloud (Abele *et al*. 2017). Therefore, various applications that (partially) use the geospatial LOD as underlying data have been fostered. One of the prominent applications fields is, akin to geospatial data in other data models, visualisation (Lemmens and Keßler, 2014).

The importance of the visualisation of geospatial data cannot be overemphasised because it allows a wide range of users to explore, synthesise, present, integrate and analyse the underlying geospatial data in an interactive manner. The Linked Data paradigm opens up opportunities for geovisualisation because it provides a mechanism of linking and consolidating distributed information. For instance, the geospatial data with coarse geometric representations can be supplemented by accurate and detailed geometries through linking with other geospatial LOD. To this end, several tools have been developed for visualising and exploiting geospatial Linked Data, e.g. Mappify (Lehmann et al. 2015), Sextant (Nikolaou et al. 2014) and Map4RDF (Llaves et al. 2014). These tools are capable of retrieving geospatial Linked Data through SPARQL endpoints and visualising the data mostly in web-based viewers. These tools, in general, use predefined styling information and render the same geometries for the features in all scales. Another notable visualisation application is the online viewer of GeoNames[3]. GeoNames is a gazetteer that plays a central role in the LOD cloud, and it generally employs one single point as the geometric representation for an entity in its dataset, and the points are clustered as the map is zoomed out to small scales. From a cartographic viewpoint, these applications are insufficient due to the lack of multiple representations in different scales for the features, which is a prerequisite for proper visualisation. This deficiency impedes the possibility of deriving appropriate geovisualisation applications from the LOD cloud.

The Linked Data paradigm has natural advantages for organising geospatial data with multiple representations as it can express the links between the representations explicitly and link representations from different sources in the web. At the meantime, driven by legislation and the open data movement, increasing number of authoritative geospatial datasets with multi-scale geometries have become open data. These factors entail the potential of enriching the LOD cloud with multi-scale geometries to foster better geovisualisation applications. Nonetheless, few studies have demonstrated the means of realising multiple representation geospatial data as Linked Data. Therefore, the goal of this paper is to pilot the publishing of multiple representation geospatial data and the linking of these data with GeoNames to accomplish adapted visualisation for this hub dataset in the LOD cloud.

2 State of the art

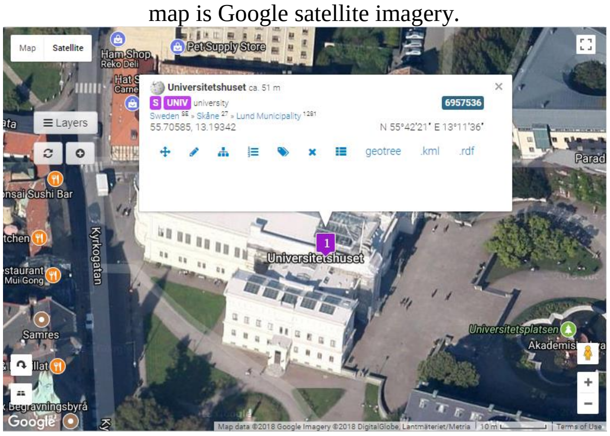
2.1 Visulisation of geospatial Linked Data

The visualisation of Linked Data, in general, refers to the techniques of visually presenting the links between entities to facilitate the intuitive discovery of underlying information and knowledge (Dadzie and Roew, 2011). For geospatial data, the

[1] http://data.ordnancesurvey.co.uk/datasets/os-linked-data
[2] http://linkedgeodata.org
[3] http://geonames.org

spatial context is crucial for easing this perception and discovery process. Therefore, the visualisation of geospatial Linked Data is often presented in the form of map mashups, in which the data are spatially represented as thematic data on top various base maps. In some of the visualisation applications, the geometric information is insufficient. Furthermore, the thematic data in such map mashups are seldom linked to the underlying data of the base map (cf. Huang *et al.* 2018). An example of these facts is the online viewer of GeoNames where e.g. the main building of Lund University (*Universitetshuset*) is visualised as a label on top of a base map (Figure 1). This visualisation technique, in some cases, could entail a significant perceptive burden for the users to understand what the feature refers to in reality; and it could also entail inferior visual performance in small visualisation scales. On the other hand, GeoNames could potentially be used for geocoding the geospatial entities in other Linked Data sources to enable geovisualisation of the data or spatial analyses. In this context, the insufficient modelling of geometry hampers the potential development of appropriate geovisualisation applications.

Figure 1: The visualization of the main building of Lund University (LU) in the online viewer of GeoNames. The base map is Google satellite imagery.



Several tools have been developed for improving the visualisation of geospatial Linked Data. For examples, LOD4WFS (Jones *et al*. 2014) was developed so that geospatial Linked Data can be retrieved through WFS protocol and visualised in GIS programs; SexTant (Nikolaou *et al.* 2013) allows visualising and browsing time-evolving geospatial Linked Data; Map4RDF (Leon *et al.* 2012) provides the possibility of editing the underlying data and selecting several types of map mashups. Nonetheless, these tools still lack the employing and handling of multiple representation geospatial data.

On the other hand, a few geospatial datasets with complex geometries have also been published as LOD. Stadler *et al.* (2012) linked GeoNames to OpenStreetMap (OSM) through the LinkedGeoData project, which opens up the possibility of using the geometric representations from OSM to visualise GeoNames through federated queries. However, the OSM, as a VGI project, generally has inaccurate and single scale geometries, which makes it still deficient for visualisation from a cartographic perspective. The complex and accurate authoritative geometries are released through the studies of e.g. Goodwin *et al*. (2009) and Vilches-Blázquez *et al*. (2014), while the multiple representation for the geospatial features are not considered.

2.2 Multiple representation for geospatial Linked Data

The publication of multiple representation geospatial data in the LOD cloud is a promising way to alleviate the flaws of the current geospatial Linked Data visualisation. The geospatial Linked Data could acquire multiple geometric representations through created links. Moreover, the Linked Data paradigm has intrinsic superiority for organising geospatial data with multiple representations. Hahmann and Burghardt (2010) compared multiple representation database (MRDB) with Linked Data, and they identified several commonalities between these two technologies. They argued that the geographic objects in both MRDB and Linked Data consist of various representations, providing a set of different views of the same object. In this context, Linked Data eases the linking of distributed representations of geospatial features. Debruyne *et al.* (2017) released authoritative geospatial Linked Data for Ireland, and the multiple representation of features is incorporated. The geometries are released in different levels of details (20, 50 and 100 metres) and stored in different named graphs. However, the semantics in terms of the visualisation scales of the geometries remains obscure.

As an increasing number of geospatial datasets has been published as Linked Data, a number of vocabularies also have been designed and released. Some of these vocabularies support multiple representations for features. For instance, the GeoSPARQL ontology defines both the concepts of *feature* and *geometry*; each feature instance can be linked to several geometry instances to enable multiple representations, and the most detailed geometry is generally defined as *default geometry* for spatial analysis. The INSPIRE directive is also investigating the solutions and potential benefits of publishing INSPIRE data as Linked Data in the ARE3NA activity[4] of the European Commission Joint Research Centre; along with this activity, several draft vocabularies for certain themes in INSPIRE (e.g. buildings and cadastral parcels) have been published[5]. Some of these INSPIRE vocabularies (e.g. the vocabularies for the theme of building) support or can be readily extended to support multiple representation. This activity derives the potential of exposing the increasing INSPIRE-compliant data into the LOD cloud, which could enrich the LOD cloud with authoritative geospatial data considerably. In the following case study, we reuse and extend the INSPIRE draft vocabularies of the theme of buildings to organise the multiple representation geospatial data; thereby from this perspective, this study is also an investigation into the potential benefits from publishing INSPIRE data as Linked Data in terms of visualisation.

3 Case study

[4] https://joinup.ec.europa.eu/collection/are3na/about

[5] https://github.com/inspire-eu-rdf/inspire-rdf-vocabularies

In this paper, we present a case study of releasing multiple representation geospatial data as Linked Data and linking to GeoNames.

3.1 Data

In the case study, we use multiple representations of the main building of Lund University (LU) as the test data. The building is located in the central part of Lund, Sweden, and it has two representations from the *Lantmäteriet*'s (the Swedish national mapping agency) multi-scale geographic datasets (Figure 2). For the visualisation of the multiple representation data, *Lantmäteriet* has recommendations for the rendering scale ranges for each level of detail. Figure 2(a) is the representation of the building in the most detailed level, it is recommended to be rendered when the rendering scale is larger than 1: 8,000; Figure 2(b) is a coarser representation that is recommended to be rendered in the scale range of 1: 8,000 to 1: 30,000. We build an MRDB (in PostgreSQL with PostGIS extenstion) through manual feature matching. However, when the test data are extended to larger area in further steps of the study, we need to employ (semi-)automatic feature matching methods (cf. Zhang *et al*. 2014), and for a complex building whose geometric representations have *n:m* (mostly *n:1* in reality) relation in the two levels of detail, the aggregations should be performed before feature matching.

3.2 Ontology design

In this case study, the ontologies are based on the INSPIRE draft vocabularies for 2D buildings, namely we mainly reuse the *bu-base* and *bu-core2D* vocabularies.

The vocabularies for buildings support buildings with multiple representations. Specifically, an instance of the *bu-core2d:Building* can be linked to several instances of *bu-base:BuildingGeometry2D* through the object property *bu-core2d:geometry2D*, in which one of the *bu-base:BuildingGeometry2D* instances needs to be specified as *reference geometry* through the datatype property *bu-base:BuildingGeometry2D.referenceGeometry*. However, the information of the rendering scale for each representation is not modelled in the vocabularies. Therefore, we develop a vocabulary to express the concepts in this respect.

The concept of scale resides in a very core place of cartography and essential for geovisualisation, and it also plays a key role in knowledge representation and measurement (Goodchild and Proctor, 1997; Carral *et al*. 2013). Therefore, the modelling of visualisation scales for the multiple representations entails substantial importance for both the visualisation applications derived from geospatial Linked Data and the sharing of cartographic knowledge. To this end, we develop a *cartographic scale* vocabulary, in which the cartographic scale is modelled as a concept (class). An object property *hasScale* is created to associate each *bu-base:Geometry2D* with a scale instance to express the information of the rendering scale. Two datatype properties are created: *hasUpperBound* and *hasLowerBound* to enable formal definition of the rendering ranges pertaining the scale instances. Figure 3 shows the core part of the INSPIRE draft building vocabularies, the cartographic scale ontology and the relations between the vocabularies.

3.3 Data transformation

The developed MRDB (including the links between different feature representations and the information of rendering scales) for the data of the main building of LU (cf. 3.1) is transformed into RDF (Linked Data) according to the adopted ontologies (cf. 3.2). The transformation is formally defined in R2RML [6] (a mapping language from relational database to RDF). The MRDB is then materialised into RDF in the triple store Stardog according to the R2RML mapping.

3.4 Interlinking with GeoNames

Interlinking with other LOD datasets adds value for the Linked Data, and it is of particular value in this study to partially justify the benefit brought from our work. We downloaded the GeoNames entities in Sweden in RDF through its search web service[7], and implemented a matching approach used in Stadler *et al.* (2012) that they employed to

Figure 2: The main building of Lund University in two levels of detail from *Lantmäteriet*.

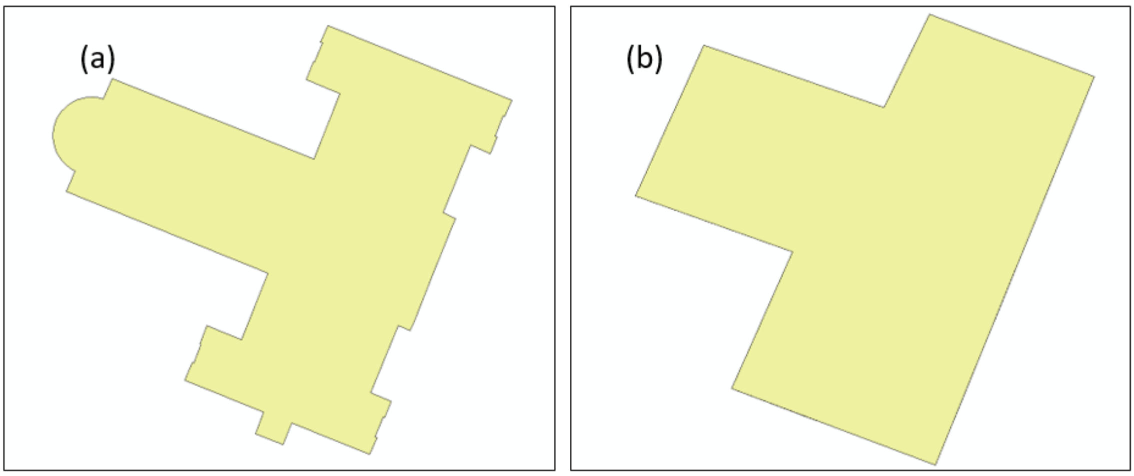
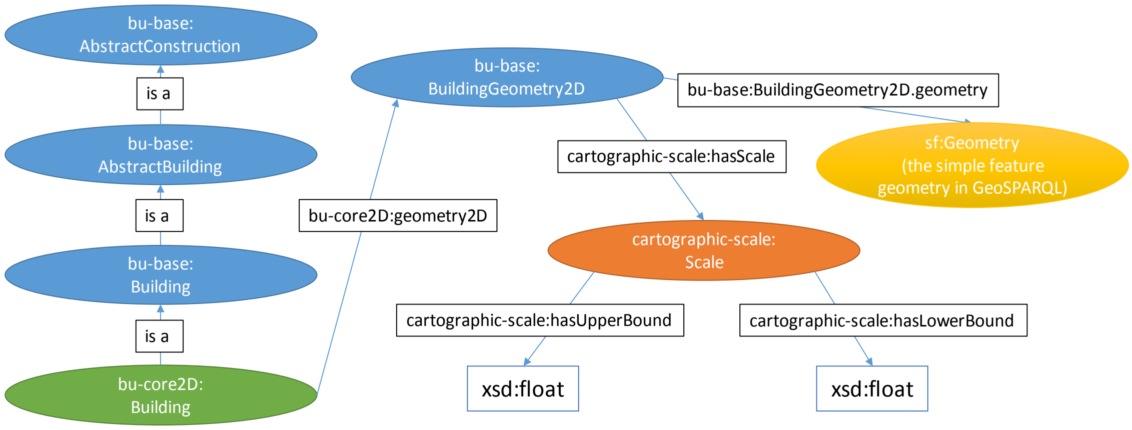


Figure 3: Illustration of the key concepts and relations in the INSPIRE draft building vocabularies and the designed cartographic scale ontology



match OpenStreetMap with GeoNames. The main building of LU in our RDF dataset is then matched to the corresponding entity in GeoNames. Since the use of the predicate *owl:sameAs* is discouraged to avoid inconsistencies (see e.g. Goodwin *et al.* (2008) for an example), we use *skos:closeMatch* (symmetric) from the SKOS vocabulary[8] to associate the instances of *bu-base:Building* and *gn:Feature* (the feature concept modelled in the GeoNames ontology[9]). The downloaded GeoNames RDF data are also stored in Stardog.

3.5 Visualisation

For the visualisation of the test data from GeoNames using the linked feature with multiple representation, we developed a web viewer. In the web viewer, the GeoNames entities are visualised using the authoritative multi-scale geometries (only the LU main building is visualised in this case).

In real-time, the RDF data are retrieved and fed to the web viewer as responses of SPARQL queries (through the SPARQL endpoint provided by Stardog). The following SPARQL query retrieves the geometries for the features in a certain rendering scale from the multiple representation geospatial Linked Data:

```
prefix gn:<http://www.geonames.org/ontology#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix bu-base: <http://inspire.ec.europa.eu/ont/bu-base#>
prefix bu-core2d: <http://inspire.ec.europa.eu/ont/bu-core2d#>
prefix geo: <http://www.opengis.net/ont/geosparql#>
prefix cartographic_scale: <[defined namespace]>

SELECT DISTINCT ?feature ?WKT
WHERE {
 ?feature a gn:Feature ;
          skos:closeMatch ?INSPIRE_building .
 ?INSPIRE_building a bu-core2d:Building ;
                  bu-core2d:geometry2D [
                     bu-base:BuildingGeometry2D.geometry [
                           geo:asWKT ?WKT
                     ];
                     cartographic_scale:hasScale [
                     cartographic_scale:hasUpperBound ?ub;
                     cartographic_scale:hasLowerBound ?lb
                  ]
               ] .
 FILTER (?ub >= [rendering scale] && ?lb < [rendering scale])
}
```

For the real-time scale-adapting visualisation, the standardised rendering pixel size is defined to be 0.28mm×0.28mm, unless the information of actual pixel size of the final display device is available. Figure 4 shows the visualisations of the LU main building in the scales of 1: 1,000 and 1: 10,000. We use the WMS service from *Lantmäteriet* as the base map and the JavaScript library Leaflet[10] to ease the development of the map viewer.

4 Work in progress

The first findings from this work provide a glance of the benefits from exposing multiple representation geospatial data in the LOD cloud. The visualisations for the test data are improved considerably compared to the single point representations in the online GeoNames viewer.

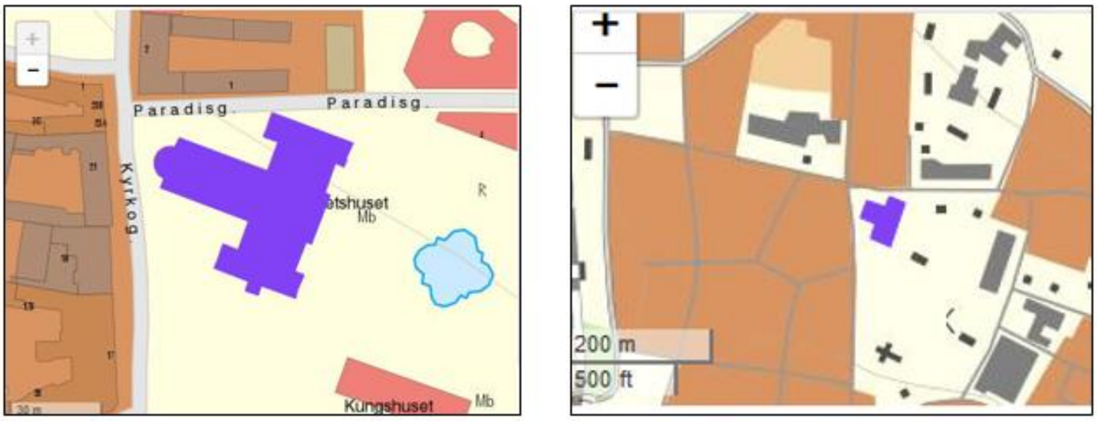
The further steps of our approach would be to enlarge the test area, which is likely to introduce some complicated cases, e.g. the buildings are aggregated irregularly or the defined rendering scale ranges do not suit some certain geovisualisation applications.

In both our study and Debruyne *et al.* (2017), the association of each feature and its multiple representations is performed in controlled environments by a mapping agency or professionals. This fashion is in line with the traditional way of creating and maintaining an MRDB. However, in the LOD environment, how to match different representations for

[8] https://www.w3.org/2009/08/skos-reference/skos.html
[9] http://www.geonames.org/ontology

[10] http://leafletjs.com

Figure 4: The visualisations of LU main building in the scales of 1: 1,000 and 1: 10,000. The base map is from *Lantmäteriet* (© *Lantmäteriet*, Dnr: I2014/00579).



features and formulate the visualisation scales still remains unclear. We argue our proposal could be used as a way for publishing authoritative geospatial Linked Data which can potentially foster better visualisation applications in a cartographic sense.

Furthermore, the Linked Data paradigm (Semantic Web technologies) could provide more benefits besides the enrichment of multi-scale geometries. The geovisualisation formal knowledge modelling using Semantic Web technologies would be a broader and more interesting topic to study, and the modelling of multiple representation and cartographic scale would be key components within that framework. Gould and Mackaness (2016) developed a knowledge graph for map generalisation using ontologies, which made a step forward concerning the knowledge modelling of geovisualisation. However, the knowledge of many other aspects of geovisualisation has not been formally modelled to facilitate the sharing of the knowledge in a both human- and computer-readable manner. We believe the advancements in this respect would substantially benefit the geovisualisation community.

Acknowledgements

The work was funded by China Scholarship Council and Lund University.

References

Abele, A., *et al.*, (2017). *Linking open data cloud diagram 2017* [online]. Available from: http://lod-cloud.net/ [accessed 15 February 2017]

Carral, D., et al., (2013). An ontology design pattern for cartographic map scaling. *In:* P. Cimiano, *et al., eds. The semantic web: semantics and big data*. Heidelberg: Springer, 76–93.

Dadzie, A.-S. and Rowe, M. (2011). Approaches to visualising linked data: A survey. *Semantic Web*, *2*(2), 89-124.

Debruyne, C., *et al*., (2017). October. Ireland's Authoritative Geospatial Linked Data. In: *The Semantic Web – ISWC 2017, Part II*. Cham: Springer, 66–74.

Goodchild, M.F. and Proctor, J., (1997). Scale in a digital geographic world. *Geographical and environmental modelling*, *1*, 5-24.

Goodwin, J., Dolbear, C. and Hart, G. (2008). "Geographical Linked Data: The administrative geography of Great Britain on the semantic web." *Transactions in GIS 12*(s1): 19-30.

Gould, N. and Mackaness, W., (2016). From taxonomies to ontologies: formalizing generalization knowledge for on-demand mapping. *Cartography and Geographic Information Science, 43*(3), 208-222.

Hahmann, S. and Burghardt, D., (2010). Linked Data-a multiple representation database at web scale?. In *Proceedings of the 13th ICA workshop on generalisation and multiple representation, 12–13 September 2010, Zürich, Switzerland*.

Huang, W., *et al.*, (2018). Synchronising Geometric Representations for Map Mashups Using Relative Positioning and Linked Data. *International Journal of Geographical Information Science*, doi: 10.1080/13658816.2018.1441416

Jones, J., *et al*., (2014). Making the Web of Data Available Via Web Feature Services. In Joaquín Huerta, Sven Schade and Carlos Granell (eds.), *Connecting a Digital Europe Through Location and Place* (Springer International Publishing), Part of the series Lecture Notes in Geoinformation and Cartography, 341–361.

Lemmens R. and Keßler C., (2014). Geo-information visualizations of linked data. In: *Proceedings of the 17th AGILE conference on geographic information science*, 3–6 June 2014, Castelln, Spain.

Leon, A.D., *et al.*, (2012). Map4rdf-Faceted Browser for Geospatial Datasets. In: *Proceedings of the First Workshop on USING OPEN DATA, W3C, 19–20 June 2012, Brussels, Belgium.*

Nikolaou, C., *et al*., (2014). Improving knowledge discovery from synthetic aperture radar images using the linked open data cloud and Sextant. In *Proceedings of ESA-EUSC-JRC 2014-9th Conference on Image Information Mining Conference: The Sentinels Era*, 63-66.

Stadler, C., *et al.*, (2012). Linkedgeodata: A core for a web of spatial open data. *Semantic Web*, 3(4), 333-354. doi: 10.3233/SW-2011-0052

Vilches-Blázquez, L.M., *et al.*, (2014). Integrating geographical information in the Linked Digital Earth. *International Journal of Digital Earth*, *7*(7), 554-575.

Zhang, X., *et al.* (2014). Data matching of building polygons at multiple map scales improved by contextual information and relaxation. *ISPRS journal of photogrammetry and remote sensing, 92*, 147-163.