Geospatial data integration and visualisation using Linked Data

Weiming Huang	Ali Mansourian	Lars Harrie
GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University	GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University	GIS Centre, Department of Physical Geography and Ecosystem Science, Lund University
Sölvegatan 12, 223 62 Lund, Sweden	Sölvegatan 12, 223 62 Lund, Sweden	Sölvegatan 12, 223 62 Lund, Sweden
weiming.huang@nateko.lu.se	ali.mansourian@nateko.lu.se	lars.harrie@nateko.lu.se

Abstract

Geospatial data are increasingly available nowadays, and this leads to more analyses and visualisation of geospatial data from several sources. To enable this, we need homogenous data as well as proper integration methods. Geospatial data integration has been a long-standing research topic for decades, and this paper discusses the utilisation of Linked Data technology stack to alleviate the geospatial data integration, particularly in the multi-scale context. Furthermore, this paper also discusses the possibilities of incorporating symbolisation information in Linked Data along with the integrated linked geospatial data for visualisation.

Keywords: geospatial data integration; multi-scale; Linked Data; visualisation; symbolisation.

1 Introduction

The rapid development of the Internet, together with the incentives from legislation, commence, and the open data trend, has led to the improvement of the availability of geospatial data, including both the authoritative geospatial data accessible from governmental Spatial Data Infrastructures (SDIs) and the prevalent Volunteered Geographic Information (VGI). For example, in Europe, the INSPIRE[1] directive formulated that in a few years' time, several authorities that are responsible for creating and maintaining geospatial data are obliged to set up download services to facilitate the access and sharing of geospatial data. The substantial improvement of data availability will enable cross-data set analysis and visualisation, in which the integration of geospatial data from different sources is indispensable.

The productions of geospatial data from different sources are generally isolated, and this causes the syntactic and semantic heterogeneity that are two significant obstacles for geospatial data integration. Furthermore, the links between multi-source geospatial data that are of relevance are often lacking. The absence of links between data sets impedes the integration of geospatial data for visualisation and analysis, and this impediment is especially significant in a multi-scale environment. For example, in a map mashup (a common form of web map), the thematic data are usually simply overlaid on the top of a base map without explicit links and integration. However, the scales of the thematic data and the base map are generally not synchronised because unlike the thematic layer, the base map is usually a multi-scale map from an authoritative mapping agency and has multiple representations (for details, see Huang *et al.*, 2016).

In this context, the Semantic Web technologies, particularly the ones concerning Linked Data, provide a promising technical framework to ease the integration and linking between geospatial data. "Linked Data" is the term for the collection of design principles and technologies centred around a paradigm to publish, retrieve, reuse, and integrate data on the Web (Kuhn *et al.* 2014). The adoption and application of Linked Data in the geospatial community have developed considerably in recent years. A number of geospatial data sets have been released as Linked Data, and some of them have made up an indispensable portion in the linking open data (LOD) cloud (The Linking Open Data cloud diagram, 2017; Figure 1). On the other hand, the visualisation and symbolisation of linked geospatial data has been rarely exploited, and it is even trickier in a multi-scale context. Hence, this project mainly concentrates on investigating the integration and visualisation of multi-source geospatial data utilising the Linked Data technology stack, in particular in a multi-scale context. The following research questions will be addressed within the work:

• How to organise geospatial data in different scales in Linked Data, the design of unique resource identifiers and ontologies is important to link the multiple representations of each geographic object;

• How to establish the links between different geospatial Linked Data sets, particularly in a multi-scale context;

• How the links between data sets can be utilised for the synchronisation of scales between multi-source geospatial data sets;

• How the linked geospatial data sets should be visualised and symbolised, namely how the symbolisation information should be defined and organised, and on which level (feature level, feature collection level, etc.) it should be defined.

• How the linked geospatial data sets would benefit the SDI.

2 Related work

2.1 Geospatial data integration using Linked Data

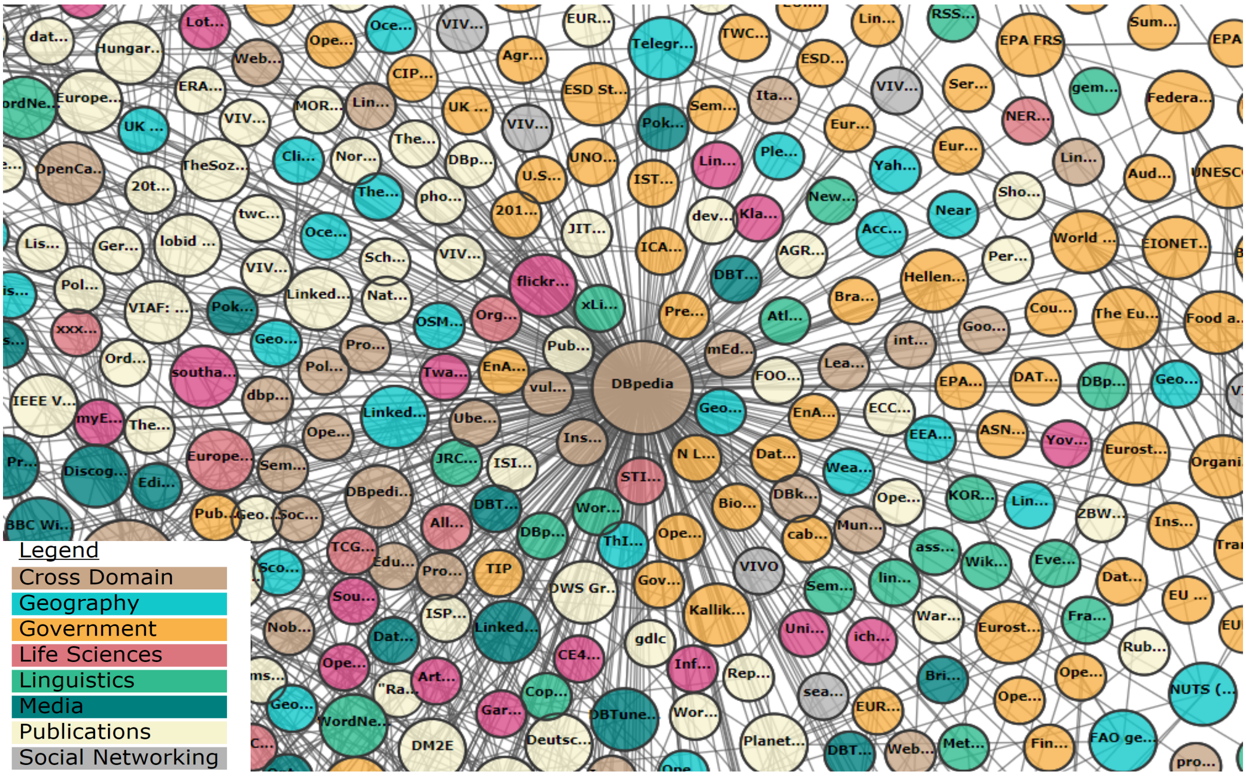

Figure 1. The central part of LOD cloud of November, 2017

Data integration is a long-standing research theme in the geospatial domain where geometric, topological as well as semantic information are used (see e.g. , Walter and Fritch 1999, Du *et al.* 2012, Yang *et al.* 2014). With a few exceptions (e.g., Mustière and Devogele. 2008), these studies have concentrated on the integration of data of similar levels of detail.

In the abovementioned environment of map mashup, in which multi-source geospatial data are generally simply overlaid together without any links established between each other, the integration usually is about multi-scale data sets. Stern and Sester (2013) studied mashups of natural protected areas on top of a base map, where the protected areas often have common geometries with the base map. To overcome the problem of inconsistencies in the multi-scale representation, they argued that the base map should act as constraints for generalising the thematic data. Toomanian *et al.* (2013) used Semantic Web technologies to integrate multi-source data in map mashups. They defined the semantic relationships between feature types in the thematic data and the base map in the map mashups using ontologies.

These semantic relationships were then used to enable real-time adjustment of the thematic features to the base map.

Linked Data technology has been adopted to facilitate geospatial data integration in some other studies. For instance, Wiemann and Bernard (2016) investigated possibilities for the integration of SDI and Linked Data paradigm in terms of spatial data integration. They implemented a prototype system where the spatial relationships were explored by the OGC Web Processing Service (WPS) and then explicitly and separately stored using Linked Data, including the information of involved features, relationship types and conducted relationship measurements. Lutz *et al*. (2009) addressed a hybrid ontology-based solution for overcoming the semantic heterogeneity in SDI. They designed a shared vocabulary on top of which the application ontologies were designed, then they used the ontology reasoner (DL query) to identify the subsumption relationships between concepts, thus the corresponding concepts in different classification systems were recognised; they also used semantic annotations to label the data services to enable the data requestor to use a tailored language to retrieve data. The tailored language was then translated into DL query and subsequently the WFS requests were invoked.

In the framework of Linked Data technology, some techniques have been extended in order to improve the handling of linked geospatial data. For example, SPARQL, as the query protocol for RDF, has a standardised geospatial extension – GeoSPARQL (Perry and Herring, 2011). GeoSPARQL also provides an ontology as a standardised exchange basis for geospatial RDF data (Battle and Kolas, 2012) and this has been adopted in several studies in which the geospatial data sets are published as Linked Data and linked to other data sets. For example, Patroumpas *et al*. (2015) exposed the INSPIRE-compliant data and metadata as Linked Data by transforming them into the data model of resource description framework (RDF) using XSLT transformations and then exposing them through (Geo)SPARQL endpoints, they adopted the GeoSPARQL ontology for the geometric representation of their RDF data sets. These technical advances enable the geospatial data to be linked and referenced. However, the linking of multi-scale geospatial data sets has been rarely explored, and this is the focus of this project.

2.2 Visualisation of geospatial linked data

The linked geospatial data are situated at rather central places in the LOD because geospatial and location data often serve as nexus and linkage between different data items and sets (Janowicz, 2012). However, the portrayal and symbolisation of linked geospatial data have been seldom discussed. When it comes to the visualisation of linked geospatial data, the providers of such data generally use external styling service or hard-coded symbolisation parameters. The LinkedGeoData (LGD) project which released OpenStreetMap (OSM) data in Linked Data used separate renderer service where the symbolisation rules are settled to render the LGD data (Stadler *et al*. 2012). The GeoNames[2] has an online portal in which the entities can be shown on the top of either a digital base map or satellite images; the entities are simply shown as labels with numerical signs or bounding boxes with uniform symbology. In these cases, the portrayal information is not explicit and can be hardly reused by the users or other organisations which are interested in the geospatial data in RDF and the visualisation of the data.

There have been some studies using ontologies to organise and semantically annotate the symbology information in Linked Data. For example, the OGC (Open Geospatial Consortium) explored semantic mediation of portrayal information of geospatial data using ontology in their testbed 11 and 12 (Fellah, 2015; 2017). They designed symbology ontologies during the testbeds, and the ontologies in testbed 11 was more inclined to the ISO 19117 standard (Kresse and Fadaie, 2004) and the ontology in testbed 12 was better aligned to Symbology Encoding (SE; Müller, 2006) and Styled Layer Descriptor (SLD; Lupp, 2007). In outline, the ontologies that they developed were modularised to avoid huge-sized ontology and foster the reusability, specifically the vocabulary was modularised into style ontology, symbol ontology, symbolizer ontology and graphic ontology. However, there still very few study concerning how the symbolisation information should be associated with geospatial information in the LOD cloud, and how the multi-scale symbolisation should be arranged if the data are in several different levels of detail.

3 Method

The Linked Data technology will be leveraged in this project. Specifically, the data will be constructed upon their connections with the reference data sets. For example, the natural protected areas are generally defined by their connections with other geographic objects (e.g., river, lake, road, etc.). Assuming that the reference geospatial data that have the topographic and cadastral objects are released in Linked Data, then the natural protected areas can be defined upon their relations with the objects in base map, and the scales between the reference data and the thematic data that are built upon the reference data can be automatically synchronised. Several case studies will be performed to verify the feasibility of the approach. In addition to this, the symbolisation information of both thematic and reference data will also be incorporated into the Linked Data sets to enable tailored visualization. The symbolisation of thematic data also can be dependent on the styling or other information in reference data.

To realise this idea, we need:

- Multiple representation databases that are released as Linked Data to serve as reference data sets, the GeoSPARQL can be employed to act as vocabulary for geometries; and the design of URI still needs to be explored;
- Ontologies that define the formal semantics of the relations between thematic and reference data, these can be extended from GeoSPARQL;
- Ontologies that define the styling information of linked geospatial data, some concepts from SE and SLD can serve as reference;
- A mechanism for generating thematic data from the relations with reference data for visualisation and analysis;
- A prototypical system that can automatically generate thematic data from the above data modelling and visualise them according to the tailored symbolisation information.

Acknowledgement

The PhD study of the Weiming Huang at GIS Centre, Lund University is jointly funded by China Scholarship Council (CSC) and Lund University.

Notes

1.https://inspire.ec.europa.eu/
2.http://www.geonames.org/

Reference

Abele, A., McCrae, J.P., Buitelaar, P., Jentzsch, A. and Cyganiak, R. (2017) Linking open data cloud diagram 2017 [online]. Available from: http://lod-cloud.net/ [accessed 15 February 2017]

Battle, R. and Kolas, D. (2012) Enabling the geospatial semantic web with parliament and GeoSPARQL. *Semantic Web,* 3(4), 355-370.

Du, H., Anand, S., Alechina, N., Morley, J., Hart, G., Leibovici, D., Jackson, M. and Ware, M. (2012) Geospatial information integration for authoritative and crowd sourced road vector data. *Transactions in GIS*, 16 (4), 455–476.

Fellah, S. (ed.) (2015) *OGC Testbed-11 Symbology Mediation Engineering Report*, Open Geospatial Consortium.

Fellah, S. (ed.) (2017) *Testbed-12 Semantic Portrayal, Registry and Mediation Engineering Report*, Open Geospatial Consortium.

Huang, W., Harrie, L. and Mansourian, A. (2016). On-demand mapping and integration of thematic data. *In ICA Workshop on Generalisation and Multiple Representation*, 14 June, Helsinki, Finland.

Janowicz, K. (2012) Place and Location on the Web of Linked Data. WWW document, http://stko.geog.ucsb.edu/location_linked_data

Kresse, W. and Fadaie, K. (2004). *ISO standards for geographic information*. Springer Science & Business Media.

Kuhn, W., Kauppinen, T., and Janowicz, K. (2014) Linked Data – a paradigm shift for geographic information science. In: Duckham, M., Pebesma, E., Stewart, K., Frank, A. ed., *Geographic information science*. Berlin: Springer, 173–186.

Lupp, M. (2007) *Styled layer descriptor profile of the web map service implementation specification*. Open Geospatial Consortium Inc. OGC.

Lutz, M., Sprado, J., Klien, E., Schubert, C. and Christ, I. (2009) Overcoming semantic heterogeneity in spatial data infrastructures. *Computers & Geosciences*, *35*(4), 739-752.

Mustière, S. and Devogele, T. (2008). Matching networks with different levels of detail. *Geoinformatica*, 12(4), 435-453.

Müller, M. (2006) *Symbology Encoding Implementation Specification.* Open Geospatial Consortium.

Patroumpas, K., Georgomanolis, N., Stratiotis, T., Alexakis, M. and Athanasiou, S. (2015) Exposing INSPIRE on the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 35, 53-62.

Perry, M. and Herring, J. (2012) *OGC GeoSPARQL – a geographic query language for RDF data* [online]. Technical report, Open Geospatial Consortium. Available from: https://portal.opengeospatial.org/files/?artifact_id=47664 [Accessed 28 August 2016]

Stadler, C., Lehmann, J., Höffner, K. and Auer, S. (2012) Linkedgeodata: A core for a web of spatial open data. Semantic Web, 3(4), 333-354.

Stern, C. and Sester, M. (2013) Deriving constraints for the integration and generalization of detailed environmental spatial data in maps of small scales. *In ICA Workshop on Generalisation and Multiple Representation*, 23–24 August Dresden, Germany.

Toomanian, A., Harrie, L., Mansourian, A. and Pilesjo, P. (2013). Automatic integration of spatial data in viewing services. *Journal of Spatial Information Science*, 2013(6), 43-58.

Walter, V. and Fritsch, D. (1999) Matching spatial data sets: a statistical approach. *International Journal of Geographical Information Science* 13(5), 445-473.

Wiemann, S. and Bernard, L. (2016) Spatial data fusion in Spatial Data Infrastructures using Linked Data. *International Journal of Geographical Information Science*, 30(4), 613-636,

Yang, B., Zhang, Y., and Lu, F. (2014) Geometric-based approach for integrating VGI POIs and road networks. *International Journal of Geographical Information Science*, *28*(1), 126-147.