On-demand mapping and integration of thematic data

Weiming Huang, Ali Mansourian and Lars Harrie

Department of Physical Geography and Ecosystem Science, Lund University

Sölvegatan 12, SE-223 62 Lund, Sweden

E-Mail: weiming.huang@nateko.lu.se; ali.mansourian@nateko.lu.se; lars.harrie@nateko.lu.se

ABSTRACT

Cartographic mashup is a powerful tool for users to visualize thematic data together with base maps, and it is increasingly popular due to the better availability of geodata. In current typical mashups, thematic data and base maps come from different providers without any connection between the data sets, and their scales are not synchronized. In this paper, we firstly explore the problems in such mashups and the potential linkages between thematic data and base maps, then propose a novel methodology to establish explicit connections between thematic data and base maps, and finally propose how to utilize them to generalise thematic data. We mainly aim at conveying our idea in this working paper; that is the paper concentrates on the feasibility of our methodology, and no implementation is presented.

Keywords: Cartographic mashups, Thematic data, Base maps, Semantic Web, RDF, Multi-scale, Linked geodata

1. INTRODUCTION

Cartographic mashups have become increasingly common for on-demand mapping. This trend is likely to continue during the next decade due to better availability of both authoritative and user generated data. A typical cartographic mashup is a thematic layer on top of base maps where the thematic data and the base maps are provided by different web based services. A main challenge to these mashups is how the scaling of the information should be done. Questions arise such as:

- Should both the base maps and the thematic data be provided in several scale ranges? And if so, should the scale ranges be synchronized between the data providers?
- Do we need to perform any integration for the data? Or is it enough that we simply add the thematic information on top of the base maps?

To synchronize the scale ranges between cartographic service providers is a difficult task. It has been discussed in e.g. the Swedish national standardisation of web cartography (TK570) but so far no recommendation is provided. Most commonly in Sweden, and to our knowledge also in other countries, the thematic data is only provided in one (or few) scale(s). We think that it is not likely that service providers of thematic data will provide their cartographic data in several scales in the near future similar to what most service providers of base maps already have done. Therefore we need methods to generalise the thematic information on-demand.

The answer to the second question is inherently dependent on the application. Most mashups today are created with thematic data that is not integrated to the base maps. Typical examples are adding point of interest (POI) and thematic polygons (cultural areas etc.) on top of the base maps. In the applications where the thematic information presents natural phenomena (e.g. geological phenomena such as rock types), this approach is preferable since there is no inherent connections between the thematic information and the base maps. But man-made/defined thematic information often has inherent connections to the base maps. It could be e.g. a natural protection area whose geographic extent is defined based on lakes shores, rivers, cadastral and other administrative borders, etc. In this study we explore if these connections could be utilized to generalize the thematic information on-demand.

It is here noteworthy to make a comparison between geodata and the data used in Building Information Model (BIM). Geodata are normally defined using absolute coordinate in a geodetic reference system, while BIM data are normally using local coordinate systems. Furthermore, it is common that the geometry of BIM data is defined based on relationships to other features so called parametric modelling (see e.g. Eastman et al. 2011). Parametric modelling is very useful for fully man-made features such as buildings, but could also be utilized for geodata (which is only partly man-made). Therefore, it could be interesting to try to utilize a similar idea to construct the geographic features within cartographic mashups, that is representing the thematic features in the form of combining several components which either come from the features on the base maps or remained from the original thematic data. On the other hand, there are currently quite a few initiatives releasing geodata in the Linked Data paradigm, e.g. the Ordnance Survey Linked Data (Ordnance Survey Linked Data, 2016) and LinkedGeoData (Stadler et al., 2012). An advantage of Linked geodata is that the relationships between the data from different layers can be represented and stored explicitly. It seems that the integration between thematic data and base maps could benefit from Linked geodata. So, one aim of this study is also to explore the usages and advantages of a type of parametric modelling as well as Linked geodata for solving the scale issue in cartographic mashups.

This working paper includes a feasibility study of creating mashups that could be used in many different scale ranges. We only study mashups where the thematic information has inherent connections to the base maps. The paper aims to answer the following questions:

- How could we generate and store connections (correspondences between the features on thematic data and base maps) between the thematic information and the base maps?
- How could we utilize information from the base maps to generalize the thematic data?

The connections between the thematic information and the base maps could be generated by (real-time) integration methods or by explicit links. This paper will concentrate on the latter approach where we propose the utilization of ontology and Linked Data technology to facilitate the connection.

The paper is organized as follows. Section two contains related work of thematic environmental data generalisationand ontologies and Linked Data techniques for cartographic data. Section three describes an example of zooming a current mashup; this example illustrates the problem that we aim to resolve by connecting the thematic information to the base maps. Then, in section four, we present a novel methodology for generating cartographic mashups. The paper ends with concluding remarks. It should be noted that this is a working paper that concentrates on the feasibility of our methodology. No implementation is done yet.

2. RELATED WORK

Sterner and Sester (2011, 2013) investigated the possibility of utilizing the features on the background maps as constraints to generalise the water protection area features on the thematic layer. They came up with an idea that during the matching process, the more often one specific landuse class is found, the more important constraint it is to preserve the topological relationships. After a statistical method, they ranked the combinations of landuse types, and utilized the possible combination with highest ranking to generalize water protection area features. In our working paper, we propose an idea to solve a similar problem as in Sterner and Sester (2011, 2013) where we utilize ontology and linked data techniques to achieve this.

An ontology can be defined as a “formal, explicit specification of a shared conceptualization” (Studer et al. 1998, p. 184). It is about clear definition of data in such a way that users and even systems have a common understanding of the semantics of data and the relationships between them. Ontologies have been frequently used to model cartographic/geographic data and we cannot make a complete review here but only provide some examples.

Lüscher et al. (2009) developed a method for ontology-driven pattern recognition. They defined an ontological model of geographic features (in their case English terraced houses). They showed how a textual description of a concept (terraced house) can be formalized into an ontology and how the ontology can be utilized to identify objects in the database (with the aim of enriching the database with semantic information).

Utilization of ontologies for data integration has been explored by Toomanian et al. (2013). They developed a method to adjust thematic vector data to fit a base map. In their case study, they developed an ontology of how historical administrative borders are related to features in modern base maps. Then this ontology was used to adjust the geometry (in real time) of the historical borders when the borders where added on top of a base map.

To utilize ontologies in web applications we need standards for storing relationships between data and between ontologies; therefore the World-Wide Web Consortium (W3C) has developed standard languages and data models such as Resource Description Framework (RDF) and Web Ontology Language (OWL). Several studies have been conducted to establish explicit linkage among geographic data or between geographic data and other data resources using RDF data model and model the semantics within the data by leveraging OWL.

The U.S. Geological Survey (USGS) piloted a process to take advantage of the Linked Data paradigm (Usery and Varanka, 2012). They developed ontologies for The National Map by combining a top-down approach based on the organization of general categories taken from standard feature classes and bottom-up approaches shaped by legacy data models, and utilized OWL format to form the vocabulary. They also converted specific datasets to RDF to make these data can be downloaded and queried in the Linked Data format. In the process of converting, they firstly converted the geographic data to Geographic Markup Language (GML) with each entity having a unique identifier which is maintained during the later converting from GML to RDF. It can further facilitate potential integration tasks.

Another Linked Geographic Data initiative was conducted in the Italian Trentino region (Shaviko et al., 2012). In order to resolve the data heterogeneity while performing various data integration tasks, they conducted an experimental work on publishing linked open geodata. They converted geo-metadata and geographic data respectively to RDF. Then they linked the RDF to some highly connected hub datasets from the linked open data cloud through using OWL's "sameAs" association. They then evaluated such experimental work in a mashup application which enables the geographic features and the external resources (e.g. images from Flickr) to be linked through Linked Data paradigm.

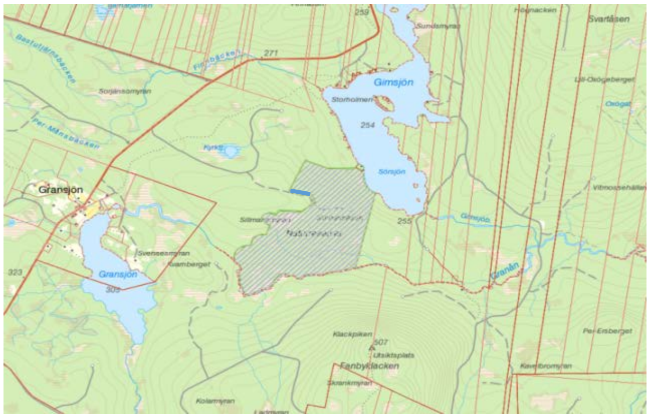
The LinkedGeoData project lifted OpenStreetMap (OSM) data into the Semantic Web infrastructure in order to simplify information integration and aggregation tasks (Stadler et al. 2012). They converted the data from OSM project to RDF and derived a lightweight ontology from OSM. After this, they performed the interlinking between LinkedGeoData and Depedia, GeoNames, the Food and Agriculture Organization of the United Nations (FOA). In order to make their project synchronized with OSM, they implemented a live-synchronization module that converts the minutely changesets published by OSM to RDF and updates such a triple set. Such project provides a promising simplified way to integrating OSM data with other data resources.

3. INVESTIGATION OF CURRENT MASHUP

In this part, we investigate a geographic mashup using the features of Swedish natural protection areas as thematic data and the cadastral units and topographic map in Sweden as base maps. In this mashup, there is no existed links between the thematic data and the base maps, so we investigate such mashup in order to identify the problems in this kind of mashup and show the potential connections between the thematic data and the base maps.

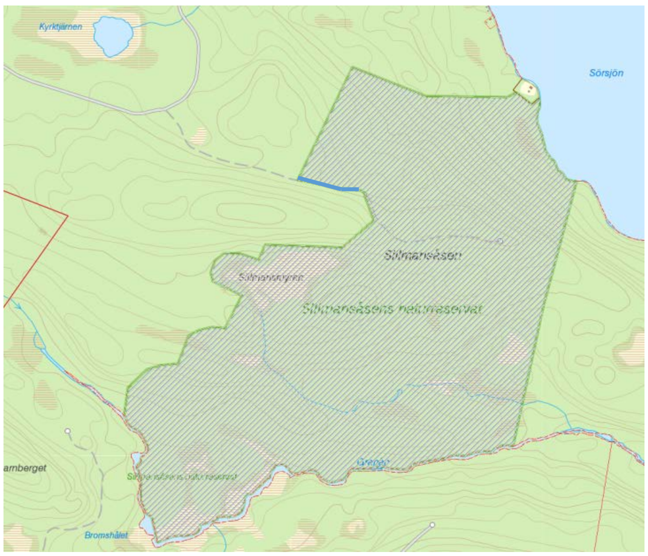
The case study of the natural protection area "Sillmansåsen" is shown in figure 1. The thematic feature is provided in one single scale. The base map of the topographic map is provided in several different scales, while the base map of the cadastral units (demonstrated by red lines on the map) is provided in one single scale.

As seen in Figure 1 large part of the border of the natural protection area coincides with the base maps (in fact, the natural protection is defined using e.g. lake shores, lanes). The natural protection areas are readable also in 1:25,000 (the smallest scale in our example), but problems are identified in larger scales. As the map is being zoomed in, that part of the feature that is connected with the lane (demonstrated by blue polylines in figure 1) becomes visually deviated from the lane after a specific scale. This is because there is no existed connection between this part of the border on the thematic map and the corresponding part of the associated lane; and partly due to that their scales are not synchronized. The aim of our idea is to resolve such a problem by creating linkage between e.g. a natural protection area and the related feature(s) elements e.g. lake shores that construct the area.



(a) 1:25,000

(b) 1:15,000



(c) 1:8,000

(d) 1:2,500

(e) 1:500

(f) 1:200

Figure 1. Illustration of the natural protection area "Sillmansåsen" in mid Sweden in various scales (zoom levels). The base maps are from Lantmäteriet (© Lantmäteriet, Dnr: I2014/00579) and the natural protection areas are from Länsstyrelsen/Naturvårdsverket.

4. PROPOSED METHODOLOGY

Our methodology comprises a pre-process and a real-time process as shown in Figure 2.

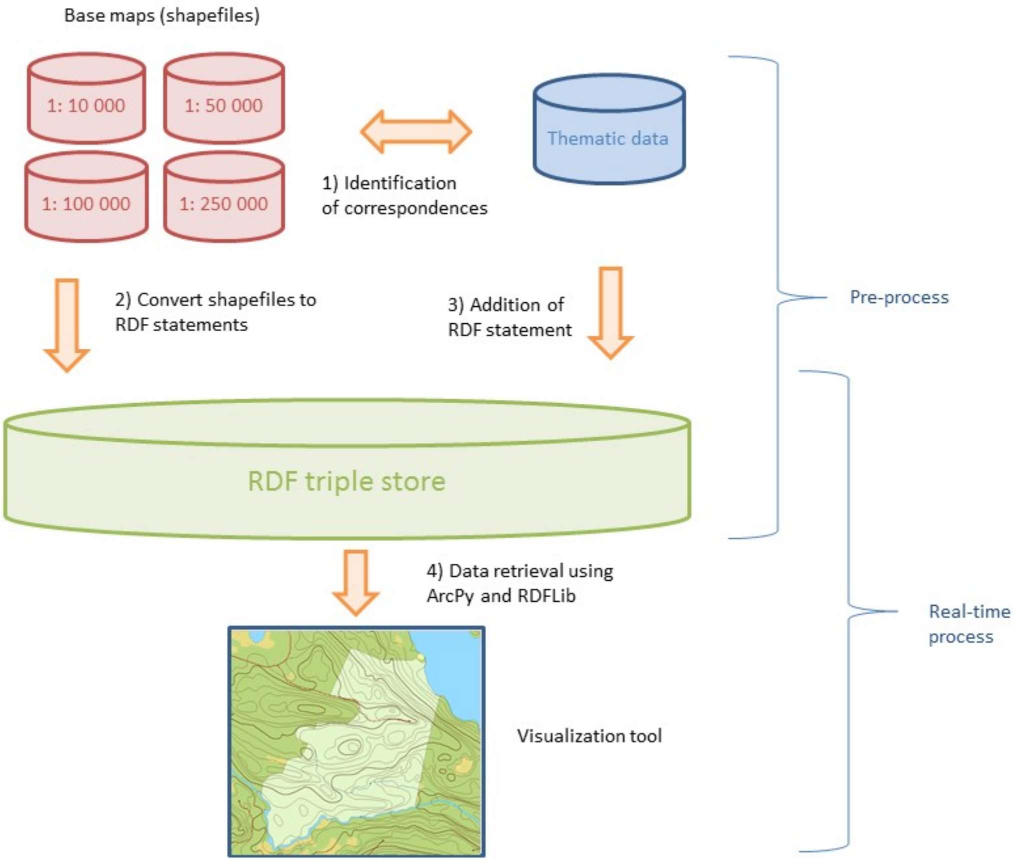


Figure 2. Flow chart of proposed methodology

1) Identification of correspondences between the thematic data and the base maps (pre-process)

The aim of this step is to identify and store relationships between the thematic data and the base maps. This could be done either using integration methods and/or text documents. For example, large part of the natural protection areas in Figure 1 could be modelled using geometries in the base maps. These relationships could either be found in text documents (the definition of the natural protection area); or more likely we could conduct matching between the natural protection area and the base maps to identify every counterpart on the base maps.

In general, a part of border of thematic data can be matched to more than one (part of) feature on the base maps, sometimes some of the counterparts are provided in several scales, and some are in one specific scale. We assign higher priority to the counterparts in several scales, and match the (part of) thematic feature to the counterpart on base maps with the highest priority. In the process of matching, the thematic features should only be matched a selection of all relevant data on the base maps where the relevance is defined if they could be used for defining the border of the thematic data. Then an automatic matching procedure is applied between the thematic data and the selected parts on the base maps. This is similar to what was done by Toomanian et al., (2013).

2) Converting the base maps to RDF statements (pre-process)

For this we plan to use the RDFLib (https://github.com/RDFLib). The RDF statements are then added to a RDF triple store (most likely Sesame, see http://rdf4j.org/).

3) Adding thematic data as RDF statements to the triple store (pre-process)

After the data matching is conducted, the correspondences that are found then will be transformed and stored by using the RDF data model along with all the geometric and attribute information of original geographic features on the base maps (the work in step two), then the thematic data under a specific zooming level will be able to represented by a combination of correspondences and original parts (for the parts that do not have any correspondences on the base maps). Here we assume that the ontologies have been defined properly using OWL.

Below we provide a simple example of which types of RDF statements that could be utilized. RDF is based on statements, where each statement consists of three parts: subject, predicate and object. In the example below we first provide some RDF statements that store the features on the base maps in three scales (in this case only one lane and one lake object). The coordinates are stored as WKT (Well Known Text) literals. Then we need statements that link the thematic feature to the base map. To accomplish this we have to divide the thematic feature into components. Finally, we have to define (again using RDF statements) which part of the base map features that are used to define the components.

……

<!--Geometric information of features on base map under several scales-->

lane_idi_1	*hasGeometry_50k*	*<coordinates>*
lane_idi_1	*hasGeometry_100k*	*<coordinates>*
lane_idi_1	*hasGeometry_200k*	*<coordinates>*

……

lake_idi_2	*hasGeometry_50k*	*<coordinates>*
lake_idi_2	*hasGeometry_100k*	*<coordinates>*
lake_idi_2	*hasGeometry_500k*	*<coordinates>*

……

<!—Representing a part of the thematic feature using the combination of parts from base map features -->

thematicFeature_idi_3	*hasComponent*	subThematicFeature_idi_3_1
subThematicFeature_idi_3_1	*isPartOf*	lane_idi_1
subThematicFeature_idi_3_1	*startsAt*	*<coordinate_of_one_point>*
subThematicFeature_idi_3_1	*endsAt*	*<coordinate_of_one_point>*
thematicFeature_idi_3	*hasComponent*	subThematicFeature_idi_3_2
subThematicFeature_idi_3_2	*isPartOf*	thematicFeature_idi_3
subThematicFeature_idi_3_2	*startsAt*	*<coordinate_of_one_point>*
subThematicFeature_idi_3_2	*endsAt*	*<coordinate_of_one_point>*

……

The approach above requires that the geometry of the thematic polygon is reconstructed all the time. This might be quite time-consuming for a real-time process. An alternative would be to create the geometries of the thematic polygons as a pre-process based on the base maps. This could be accomplished by using an ontology to define the relationships between the thematic polygon and the base maps in conjunction with an integration method. An outcome of that would be the geometry of the thematic polygon in each scale range. Then all of these geometries has to be stored as RDF statements. The drawback with this is of course that you get multiple storage of the coordinates which could lead to inconsistencies when the base map is updated and the idea of using a parametric representation is somewhat lost.

4) Retrieving, manipulating and visualizing data (real-time process)

The real-time process is triggered when a user visualizes the thematic information on top of base maps, e.g. creates a mashup. Firstly, the base maps are rendered at the specified visual scale (zooming level). Then the visual geometry of the thematic features is constructed from the parts of thematic features with the counterparts on the base maps, and maintaining the parts of thematic features that do not have any counterpart on the base maps. In this way, thematic features will have the same geometric presentation as their counterpart features in the base maps have, on a particular scale. A challenge is here to define the different parts of this process as an ontology (cf. Gould and Mackaness, 2016) and utilize this ontology in the real-time process.

The data retrieving will be conducted by querying the triple store using SPARQL (using RDFLib), and then the manipulation and visualization of the retrieved data will be performed in a tool developed using the ESRI Python programming interface ArcPy.

5. CONCLUDING REMARKS

Linking thematic data with base maps in cartographic mashups by leveraging Semantic Web technologies can not only improve the performance on the readable aspect and make the information conveyed from the mashups more meaningful, but also lift the geodata into the Semantic Web infrastructure. It can facilitate further data integration beyond the narrow scope of a specific cartographic mashup. Providing that in our methodology, the correspondences between thematic data and base maps are explicit and have been stored within the Linked Data paradigm, it should be less time-consuming than the real-time matching tasks during the visualization process. We could also note that the geometric representations used for the thematic polygons are based on features in the base map, an idea borrowed from parametric representations used in e.g. BIM.

However, the implementation of our proposal is challenging. It is a complicated process, and some procedures within it could be very subtle, e.g. which algorithm to be utilized to conduct matching in the pre-process, is the real-time process still too time-consuming for the users, etc. In this paper, we proposed some general ideas and the next step is to implement and evaluate the ideas.

Acknowledgements

We would like to express our thankfulness to China Scholarship Council (CSC) for supporting the first author's PhD study at Lund Univeristy. Financing of second and third author from Lund University are acknowledged. We also thank Lantmäteriet and Länsstyrelsen/Naturvårdsverket for proving geodata for our case study.

References

Dorn, H., Törnros, T. and Zipf, A., 2015. Quality Evaluation of VGI Using Authoritative Data—A Comparison with Land Use Data in Southern Germany. *ISPRS International Journal of Geo-Information*, 4(3), pp.1657-1671. DOI: 10.3390/ijgi4031657

Eastman, C., Eastman, C.M., Teicholz, P. and Sacks, R., 2011. *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*. John Wiley & Sons.

Gould, N. & Mackaness, W. 2015. From taxonomies to ontologies: formalizing generalization knowledge for on-demand mapping. *Cartography and Geographic Information Science***,** 1-15.

Lüscher, P., Weibel, R., & Burghardt, D., 2009. Integrating ontological modelling and Bayesian inference for pattern classification in topographic vector data. *Computers,Environment and Urban Systems, 33*(5), 363–374.

Ordnance Survey Linked Data, 2016. Ordnance Survey, Accessed 10 May 2016, http://data.ordnancesurvey.co.uk/datasets/os-linked-data

Shvaiko, P., Farazi, F., Maltese, V., Ivanyukovich, A., Rizzi, V., Ferrari, D. and Ucelli, G., 2012. Trentino government linked open geo-data: A case study. In *The Semantic Web–ISWC 2012* (pp. 196-211). Springer Berlin Heidelberg.

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

Stern, C. and Sester, M., 2011. Towards extraction of constraints for integrating environmental spatial data in digital landscape models of lower resolution–a work in progress. In *14th ICA/ISPRS Workshop on Generalisation and Multiple Representation, Paris*.

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 Map Production, Dresden, Germany*.

Studer, R., Benjamins, V.R. and Fensel, D., 1998. Knowledge engineering: principles and methods. *Data & knowledge engineering*, *25*(1), pp.161-197.

Toomanian, A., Harrie, L., Mansourian, A., Pilesjö, P., 2013. Automatic integration of spatial data in viewing services, *Journal of Spatial Information Science*, Num. 6, pp. 43-58, doi:10.5311/JOSIS.2013.6.87.

Usery, E. L., Varanka, D. (2012). Design and development of linked data from the national map. *Semantic Web, 3*(4), 371-384.