Towards Linked Data and ontology development for the semantic enrichment of volunteered geo-information

Rob Lemmens
University of Twente,
Faculty of Geo-Information Science and Earth Observation (ITC)
Enschede, The Netherlands
r.l.g.lemmens@utwente.nl

Gilles Falquet, Claudine Métral
University of Geneva, Faculty of economics and social sciences,
Geneva, Switzerland
{Gilles.Falquet; claudine.metral}@unige.ch

Abstract

The increasing volume of Volunteered Geo-information (VGI) often remains hidden and unconnected to other sources of information, and therefore hampering its reuse. A better understanding of the semantics behind VGI and its resources can relieve this issue, facilitated by the use of lightweight Semantic Web tools for the visualization, exploration and querying of Linked Data. The support of the publication of VGI-system descriptions on the Web eases the reuse of VGI, which in its turn supports the demand for a better spatial coverage of volunteered geo-information. This paper describes an approach for the collaborative development of an ontology of VGI concepts and the creation of Linked Data for real-life VGI resources. A corpus of VGI concepts and resources are drawn from the ENERGIC project. The paper demonstrates how these resources can be integrally accessed through a visual query environment.
Keywords: Volunteered Geo-information (VGI), Ontology design, Linked Data, Semantic enrichment

1 Introduction

Volunteered Geo-information (VGI) can be valuable to complement authoritative data held in Spatial Data Infrastructures (SDIs). Although there are many VGI-producing projects, there is, apart from the larger initiatives such as OpenStreetMap, not much attention for the reuse of the volunteered information beyond the project that has produced this information. We believe that the disclosure of the right meta-information serves the first step to achieve this. In this paper we describe how we create and use an ontology of VGI concepts within the ENERGIC project as a basis to semantically enrich this meta-information.

The ENERGIC project (European Network Exploring Research into Geospatial Information Crowdsourcing: software and methodologies for harnessing geographic information from the crowd; COST Action IC1203)[1] aims at finding new solutions for the exploitation, integration and application of user-generated geo-information. The project work is performed along three cross-cooperating working groups: (1) Societal and human aspects of VGI, (2) Spatial data Quality and infrastructures and (3) Data mining, semantics and VGI. ENERGIC aims to deliver an open and updatable repository of VGI analysis and integration tools and methods, literature and case studies.

The ontology which serves as semantic backbone for the repository has the following objectives:

1. Common understanding of VGI concepts (e.g., for indexing literature, outreach and education) – delimiting the field.
2. Create applications and database schema for the development of new VGI applications.
3. Usage of ontology of tasks for evaluating of data quality.
4. Semantic enrichment of VGI-based systems and indexing publication for VGI source discovery and integration.
5. Create relationships between VGI and other domains, such as GIS, SDI, etc.

In the process of building an ontology of VGI concepts, we used the draft text of the forthcoming ENERGIC book [3] as a basis for a corpus of terminology.

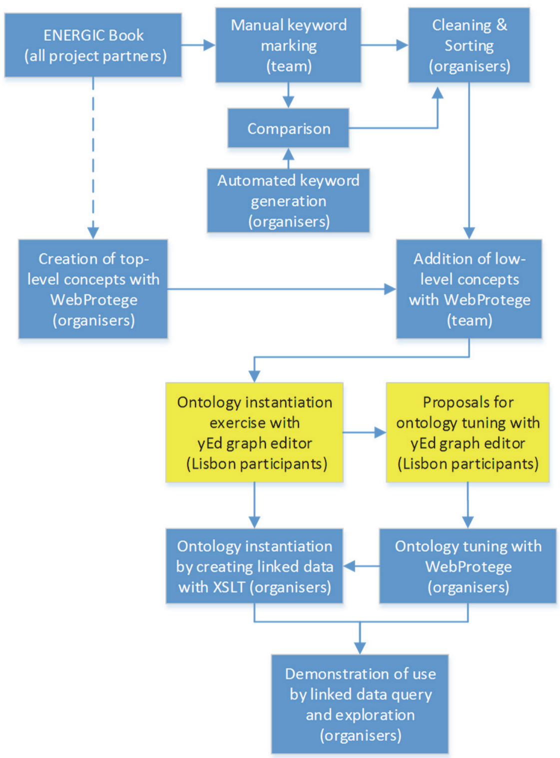
2 Semantic Web

2.1 Ontology engineering

Methods for creating ontologies have been reported in literature over the last decades, highlighting different aspects, such as formalization [10], co-creation [4,9], reuse [8] and geo-information specific issues [2,6]. In our effort we strive for a collaborative design in order to meet the abovementioned community-driven objectives. The overall process is depicted in Figure 1. The ENERGIC book [3] is considered to contain the relevant terms used in the VGI domain. The collaborative aspects of the process of ontology creation are explained in Section 3. The use of the ontology is described in Section 4.

[1] http://vgibox.eu

Figure 1: Workflow executed to obtain the ENERGIC VGI ontology. The authors of this paper are the organisers of the workflow, assisted by a small team of project partners, knowledgeable of concept modelling. The 'Lisbon participants' constitute a group of project partners involved in the use of VGI and participating in a project workshop (for a further explanation, see Section 3).



2.2 Linked data

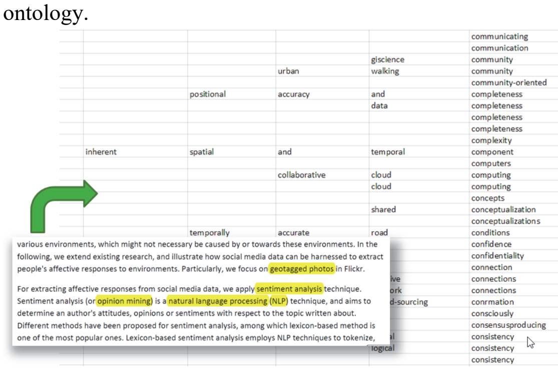
Besides expressing conceptual models with the help of ontologies and representing them in formal ontology languages such as OWL, we can make data more meaningful by simply linking their representations on the web. The Resource Description Framework (RDF) allows us to define relationships (predicates) between any two data elements (subject and object), known as triples. Sets of triples used to represent data on the Web are known as Linked Data [5]. Linked Data allows us to find data on the Web and to query across different data sources and to do this in conjunction with web-based ontologies through the SPARQL query language. The ENERGIC VGI ontology is written in OWL, which allows both a visual representation for user interaction and machine interaction for query operations together with Linked Data representations of VGI.

3 Collaborative Development of VGI Ontology

3.1 Concept extraction from keywords

The first step in the creation of the ontology was carried out by the authors with help of a small project team and involved the manual marking of keywords in the 40+ pdf documents which make up the (late 2015) draft the ENERGIC book (the final version is not expected to differ much in terms of keywords). The extracted 2634 phrases were cleaned from duplicates and unwanted characters, which resulted in 1977 keyword phrases (containing multiple keywords). The phrases were sorted on the last keyword and were published for the team to serve as a list of candidate ontology concepts. A schematic view of the extraction process is depicted in Figure 2. A comparison was done with a few automatic procedures for keyword generation, but this revealed that the manual method produced more valuable keywords.

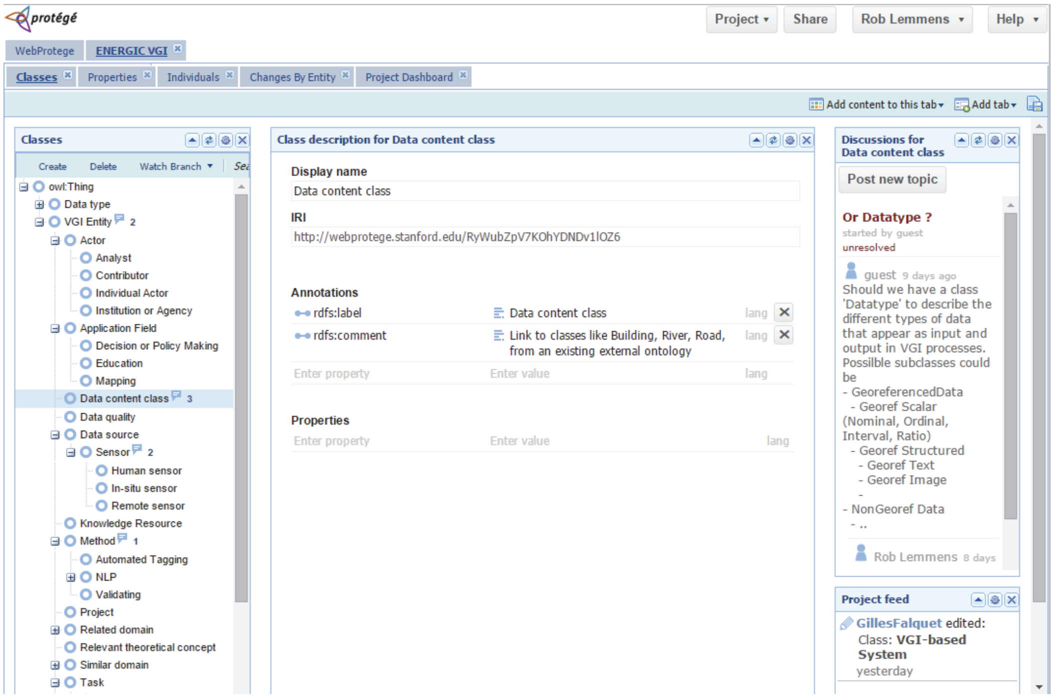
Figure 2: Keyword extraction, serving as input to VGI ontology.



3.2 Concept relationships

After the initial extraction and combination of terms into a machine-readable ontology, it was evaluated by the project team and was followed by building a framework of concepts and their interrelationships. The team was asked to add keywords from the generated list into the ontology. For this we used Web Protégé as a collaborative tool. Figure 3 shows the ontology (under construction) in Web Protégé. Although the ontology was initially organised solely as a hierarchy of classes, it is now being populated with object properties.

Figure 3: Web Protégé online ontology editor with ENERGIC VGI concepts. The panel on the right accommodates discussions on edits by the collaborators.



3.3 Lisbon workshop

On November, 19th, 2015 a workshop was organised in Lisbon to allow for input from 40 project partners. During this workshop the participants were asked to use the ontology to describe examples of VGI-based systems and literature for two reasons:

1. To create Linked Data for the purpose of semantic enrichment (see Section 4.1)
2. To identify gaps and inconsistencies in the ENERGIC VGI ontology.

The participants were assumed to have knowledge of VGI, but minor or no knowledge about ontology creation. Actually, the ten team members assisted the participants during the assignments with the tools at hand (see Section 4.1). The 40 participants worked in 8 groups and created in 1.5 hour several ontology-based descriptions of VGI-based systems and literature. They used 15 concepts of the existing VGI ontology (Figure 3) and suggested 55 new concepts. Several of these terms have been added to the ontology after the workshop.

The results of the Lisbon workshop were:

- The participants were generally able to describe systems with the proposed vocabulary.
- The concepts were sound and understandable. This can be concluded from the provided descriptions, being consistent with the semantics of the used ontology concepts.
- The coverage of basic concepts was good, but it appeared that a few more detailed aspects were missing (in particular the topic of gamification in volunteering geo-information, data quality and multi-media inputs)

4 Deployment with VGI resources

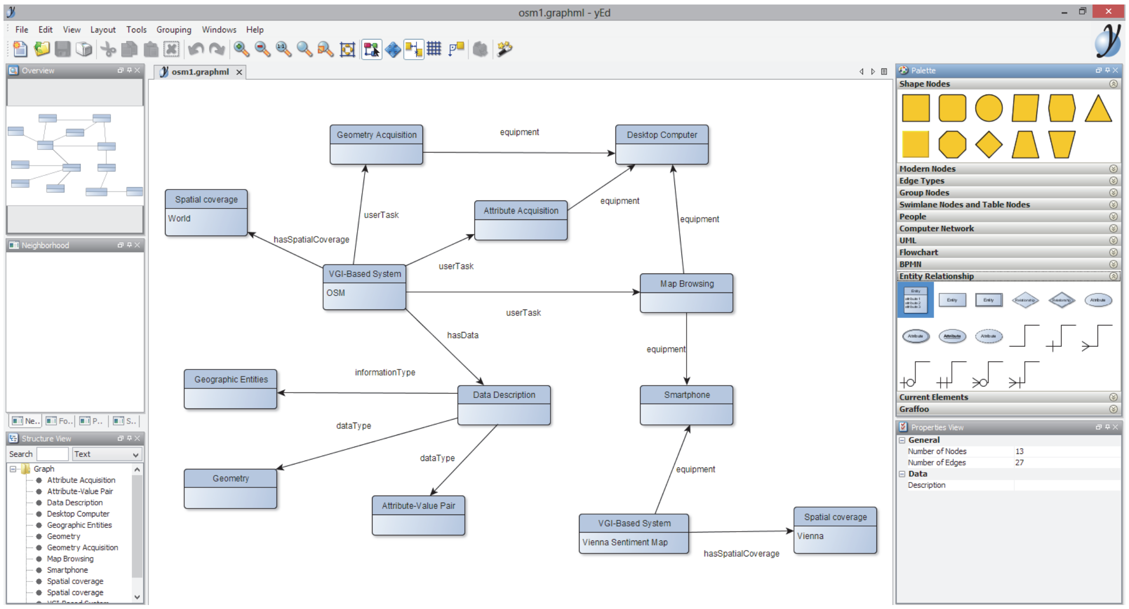
The deployment of the ontology manifests itself by being able to describe VGI resources with it and to integrally access these descriptions for information discovery, integration, exploration, etc. (see the objectives listed in Section 1).

4.1 Semantic enrichment

Linked Data enables us to combine the semantics of data sets to enrich one or both. Illustrative examples of a posteriori enrichment are provided in [1] by linking data of OpenStreetMap (spatially-rich and semantically-poor) and DBPedia (semantically-rich and spatially-poor) and in [7] by aligning a domain ontology with upper level ontologies (generic knowledge) to enrich bibliographic databases.

In our project we opted to involve VGI users to manually instantiate the VGI ontology classes with VGI resource instances. The graph editor *yEd* which was used, allows these users to create VGI instances and visually connect them to ontology classes. The graph editor stores all connections in graphml (XML). A small Java program was created to convert graphml into RDF to make the enriched VGI meta information ready for querying. Figure 4 shows a user's model of OpenStreetMap (OSM) and a representation of the Vienna Sentiment Map project (one of the projects reported in [3]).

Figure 4: Instantiation with yEd graph editor.



4.2 Exploratory querying

Several, though not an abundant number of, tools allow for the visualization of ontology and Linked Data structures (e.g., Gruff[2] and WebVowl[3]). Apart from the sole visualization of Linked Data-based VGI descriptions for the purpose of exploring and understanding its structure, there is a need for targeted queries when it comes to data integration and data retrieval in end-user applications. A way to combine best of both worlds is known as *exploratory querying*. [11] describes this approach in a spatio-temporal context and provides a tool called SPEX, which allows users to visually construct a query on Linked Data with the help of spatio-temporal constraints. Figure 5 shows the user interface in which a query is executed on the VGI as depicted in Figure 4. As exploration and querying go hand in hand, the user is able to fine-tune the query in an interactive manner and get to know the semantics of the VGI resource(s) at hand. Examples of queries that can be answered with this ontology are e.g., "Find the platforms that collect air quality data in Southern Europe", "Find articles that provide results on User participation patterns", ...

[2] http://allegrograph.com/gruff2/

[3] http://vowl.visualdataweb.org/webvowl.html

5 Conclusions and Future work

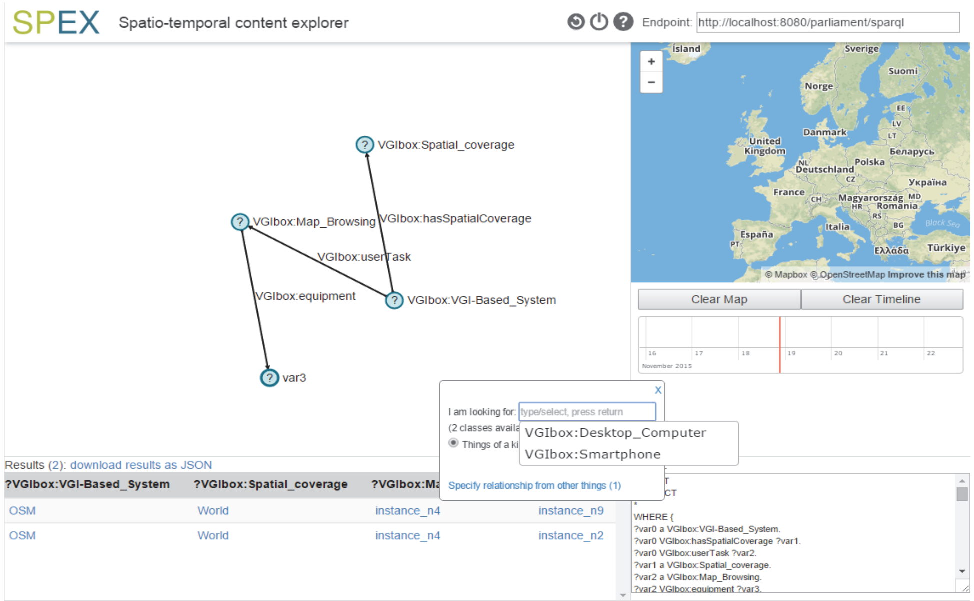
As the VGI Ontology is still under construction – and it will stay as such for a while – it already appears to be a valuable mechanism for a common understanding of the VGI domain. In this early stage of the deployment we have seen that describing VGI resources is feasible for users in a wider context, i.e., for literature, data and methods, with minor to no knowledge of conceptual modelling required. Basically, by testing the applicability of the ontology in the Lisbon workshop as described in Section 3.3, we have started to address objectives 1 and 4 as described in Section 1.

Future work will embark upon more usages of the descriptions created with the help of the ontology. Those descriptions are likely to be published as linked data (using the abovementioned RDF-triples) in the Linked Open Data cloud, which provides opportunities to connect the VGI domain with other domains, such as GIS, DBPedia and Geonames.

As immediate follow-up action the ENERGIC project has started to evaluate additional tools such as Semantic Media Wiki, for further populating the project repository and making use of the VGI ontology. In addition, since we have only processed a limited number in our list of highlighted keywords, we need to continue this by consulting the VGI community.

To further develop strategies in deploying Linked Data and to come up with innovative approaches, the project is

Figure 5: Exploratory querying Linked Data with the SPEX tool. The interface contains a visual query window (top-left), a spatio-temporal condition window (top-right), query result window (bottom-left) and (Geo)SPARQL code window (bottom-right).



organizing a multi-location and multi-event Datathon[4], challenging participants to connect existing VGI resources through the Semantic Web and compare this with conventional ways to mashup VGI, social media content and authoritative data.

References

[1] Ballatore, A., & Bertolotto, M. (2011). Semantically enriching VGI in support of implicit feedback analysis. Lecture Notes in Computer Science, 6574 LNCS, 78–93. doi:10.1007/978-3-642-19173-2_8

[2] Buccella, A., Cechich, A., Gendarmi, D., Lanubile, F., Semeraro, G., & Colagrossi, A. (2011). Building a global normalized ontology for integrating geographic data sources. Computers and Geosciences, 37(7), 893–916. doi:10.1016/j.cageo.2011.02.022

[3] Capineri, C. et al. (Eds.) European handbook of crowdsourced geographic information. Ubiquity Press, Forthcoming.

[4] Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. AAAI-97 Spring Symposium Series, SS-97-06, 33–40. doi:10.1109/AXMEDIS.2007.19

[5] Heath, T., & Bizer, C. (2011). Linked data: Evolving the Web into a global data space (1st edition). Synthesis Lectures on the Semantic Web Theory and Technology 11 (Vol. 1). doi:10.2200/ S00334ED1V01 Y201102WBE001.

[6] Janowicz, K. (2012). Observation-Driven Geo-Ontology Engineering. Transactions in GIS, 16(3), 351–374. doi:10.1111/j.1467-9671.2012.01342.x

[7] Lacasta, J., Nogueras-Iso, J., Falquet, G., Teller, J., & Zarazaga-Soria, F. J. (2013). Design and evaluation of a semantic enrichment process for bibliographic databases. Data and Knowledge Engineering, 88, 94–107. doi:10.1016/j.datak.2013.10.001

[8] Leung, N. K., Lau, S. K., Fan, J., & Tsang, N. (2011). An integration-oriented ontology development methodology to reuse existing ontologies in an ontology development process. Proceedings of the 13th International Conference on Information Integration and Web-Based Applications and Services, 174–181. doi:10.1145/2095536.2095567.

[9] Nicola, A., & Missikoff, M. (2016). Methodology for Rapid Ontology Engineering. Communications of the ACM, 79–86.

[10] Rajpathak, D., & Chougule, R. (2011). A generic ontology development framework for data integration and decision support in a distributed environment. International Journal of Computer Integrated Manufacturing, 24(2), 154–170. doi:10.1080/0951192X.2010.531291

[11] Scheider, S., Degbelo, A., Lemmens, R., van Elzakker, C., Zimmerhof, P., Kostic, N., … Banhatti, G. (2015). Exploratory querying of SPARQL endpoints in space and time, 0(0), 22. Retrieved from http://www.semantic-web-journal.net/system/files/swj1163.pdf

[4] http://vgibox.eu/activities/datathon-challenge/