Short Paper: An Ontology Design Pattern for Spatial Data Quality Characterization in the Semantic Sensor Web

Auriol Degbelo

Institute for Geoinformatics, University of Muenster, Muenster, Germany
degbelo@uni-muenster.de

Abstract. Quality is an important aspect of data discovery in the Semantic Sensor Web. This work extends current endeavors to make the Sensor Web more semantic by introducing an ontology design pattern which facilitates the modeling of aspects of spatial data quality. The implementation of a software program over two scenarios demonstrates the usefulness of the ontology design pattern for the Semantic Sensor Web.

1 Introduction

The Semantic Sensor Web (SSW) is defined after [12] as a framework for providing enhanced meaning for sensor observations and enabling the awareness of the situations that sensors observe. The SSW complements the efforts of the Sensor Web Enablement (SWE) Initiative of the Open Geospatial Consortium (OGC) [1]. The main technologies of the SSW are OGC standards (e.g. the Sensor Observation Service), ontologies (e.g. the Semantic Sensor Network Ontology) and rules (e.g. the Semantic Web Rule Language).

A review of the recent developments of the Sensor Web Enablement framework of the OGC was done in [3], and it identified the following open challenge:

> "Knowledge about the quality ... of sensor outputs is essential for making the right decisions based upon observations. At the moment, such information is often missing in observations and there is no unique way of how to incorporate it".

This work suggests an ontology design pattern [2] (ODP) to describe quality aspects of sensor observations [3]. It builds upon ontologies for sensors and observations presented in [9,4] and aims at complementing them. Section 2 presents a brief introduction to spatial data quality components. Section 3 introduces the ODP to describe quality aspects, section 4 discusses the validation of the ODP and section 5 concludes the paper.

---

[1] See an introduction to the OGC SWE in [2].

[2] An ontology design pattern is defined after Gangemi and Presutti [6] as a modeling solution to solve a recurrent design problem. Ontology design pattern as used here refers to *content ontology design pattern* (see [6] for details).

[3] Throughout the paper the terms 'sensor observation', 'observation', 'data source', 'data' and 'dataset' are used interchangeably.

2 Quality and components of spatial data quality

Two major aspects of quality characterization in the SSW can be distinguished: quality aspects that have to do with the *characteristics of the data sources*, and those that have to do with *the creation of applications based on these data sources*. These two categories of quality aspects were already summarily mentioned in [5]. The current work discusses only quality aspects of data sources. In particular, it focuses on *spatial* data quality, in view of the fact that the SSW reposes on standards developed by the Open *Geospatial* Consortium. The following definition (reflecting a data consumer perspective) to the term is suggested. It is adapted from [13].

> *Quality is the degree to which a data or service fulfills the needs of a consumer. It is a function of intangible properties (of the data) considered pertinent to the satisfaction of the consumer's needs.*

The intangible properties considered pertinent to the satisfaction of the consumer's needs are also called *components of data quality*. As regards spatial data, quality components vary from author to author. For instance, ISO 19113 includes completeness, logical consistency, positional accuracy, temporal accuracy and attribute accuracy (see [10]); Paradis and Beard [11] includes accuracy, resolution, consistency and lineage. This section will not review all the components of spatial data quality. Instead, for the purposes of the illustration (see section 4), the quality component 'resolution' is chosen. *Resolution* is defined here as the amount of detail in the dataset. Spatial resolution of raster data can be measured using the size of the raster cells; for spatial resolution of vector data, several measures are possible (see [7] for the discussion).

3 ODP for spatial data quality characterization

Gangemi and Presutti [6] suggest four ways of creating content ontology design patterns (CPs), namely: (i) reengineering from other data models, (ii) specialization/composition of other CPs, (iii) extraction from reference ontologies, and (iv) creation by combining extraction, specialization, generalization, and expansion.

The pattern presented in this section is obtained by *extraction from the Stimulus Sensor Observation (SSO) ontology design pattern* [4]. Two classes are left out: the stimulus and the sensor. They are not included because for a quality assessment operation, there is no need to describe the stimulus, nor is there a need to describe a sensor that performs a measurement. Instead, it suffices that the data consumer describes the *procedure used for quality assessment*.

The documentation of the pattern uses the fields suggested in [6]. A small difference though, is that the field **Diagram** points to a conceptual map depicting the ODP aligned to the foundational ontology Dolce Ultra Light (DUL), instead of a UML class diagram as initially suggested in [6]. The pattern is encoded in the Web Ontology Language (OWL) using Protégé [5].

[4] or the Semantic Sensor Network (SSN) Ontology.

[5] `http://protege.stanford.edu/`.

Name: ontology design pattern for spatial data quality characterization

Intent: to describe spatial data quality components of sensor observations

Competency questions: (i) what is the value of the spatial data quality component for this sensor observation? (ii) what is the quality criterion used to assess the quality of the sensor observation?

Elements: The pattern has 5 elements: Data, DataQualityCriterion, DataQualityComponent, DataQualityObservation and DataQualityResult. A *data* is the output of an observation process involving a sensor, a stimulus, a sensed property and a feature. It is equivalent to 'Observation' as defined in the SSN ontology. A *DataQualityComponent* is any property of the data that the consumer would like to approximate. Examples of spatial data quality components were mentioned in section 2. A *DataQualityCriterion* is a criterion defined by the data consumer to get information about the quality of the data. A *DataQualityObservation* is an operation by which a data quality value is assigned to a data quality component using a data quality criterion. The outcome of a data quality observation is a *DataQualityResult*.

Diagram: see Figure 1

Consequences: *Benefits:* (i) Reasoning and inference of spatial data quality component values for existing sensor observations (ii) Detection of inconsistencies during the integration of observations with different quality levels (iii) Detection of inconsistencies during the integration of observations for which different quality criterion have been used to assess the spatial data quality. *Trade-offs:* The pattern does not give a quality value like 'high quality' or 'low quality' as an end result. Instead, it helps to infer the value of a spatial data component (e.g. resolution = 20m) and it is left up to the data consumer to decide whether 'resolution = 20m' means high or low quality for the task at hand.

Known uses: see examples of uses in section 4

Extracted from: the SSO ontology design pattern / the SSN ontology

Building block: `http://wsmls.googlecode.com/svn/trunk/global/Patterns/Quality/DataQuality/dataqualitymodule.owl`

4 Validation of the ODP

A software program was developed to test the usefulness of the ODP. This software program serves as a proof that (i) the pattern can be used to perform inference of quality component values and (ii) the pattern can be used to warn against the integration of datasets for which different quality criteria are used for quality assessment. This method for validation falls into the category 'Empirical validation' introduced in [8]. Regarding the technologies, Java was used as programming language, inference rules were written using the Semantic Web Rule Language (SWRL), Pellet was used as reasoner, the ODP was accessed using the OWL API, and Jena was used to perform SPARQL queries over the ontology. Ontology instances were added to the ODP in order to answer the

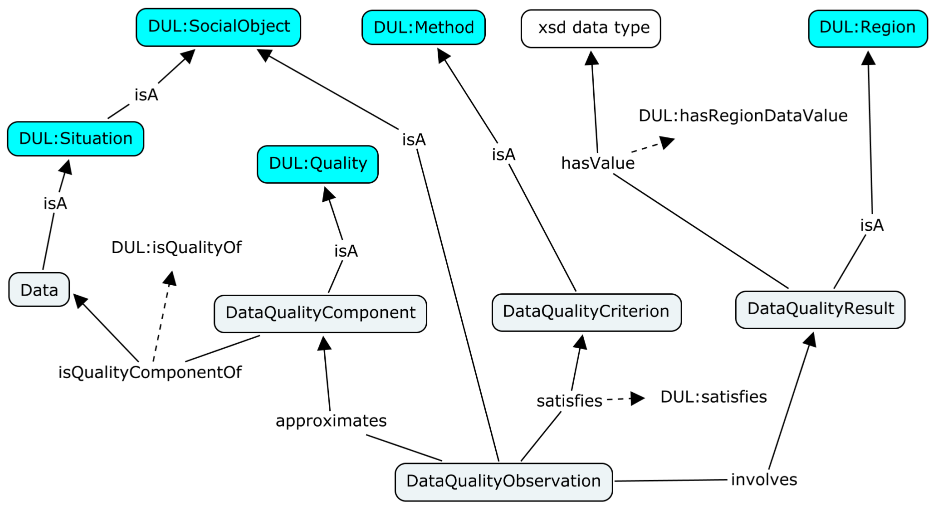

Fig. 1. Ontology Design Pattern for spatial data quality characterization aligned to Dolce Ultra Light (the prefix DUL indicates concepts from Dolce Ultra Light). The alignment of the pattern to Dolce Ultra Light is inherited from the alignment of the SSO to Dolce Ultra Light.

two questions posed in ScenarioA and ScenarioB. The Java code is available at `http://ifgi.uni-muenster.de/~degbelo/Resources/SSN2012Degbelo` or upon request to the author.

ScenarioA - *Inference of data quality value.* A decision-maker has at his/her disposal different OGC services delivering data about an observed property. He/she would like to answer the question - what is the value of the spatial resolution for an observation offering?[6] - in order to compare them and select the most appropriate for his/her task. A quality criterion for the observation offering is the sampling density[7] (i.e. number of sensors per square meters).
Comment: Information about the spatial resolution of an observation offering can be deduced from the GetCapabilities file of an OGC service. Using the ODP introduced earlier for this scenario, it is only required to parse the GetCapabilities file and assert (or store) the spatial extent of the observation offering as well as its number of sensors. The spatial resolution can then be inferred using a SWRL rule (see details in the Java code). Inference in turn is useful to address one of the drawbacks of adding semantic annotations to sensor nodes in sensor networks. In effect, [1] pointed out that adding semantics to sensor nodes in a

[6] The question is an application of the competency question (i) from section 3 to the spatial data quality component 'Resolution'.
[7] This is only one way of assessing the resolution of an observation offering. Resolution as defined in the SSN ontology can be inferred from the characteristics of the sensor.

sensor networks implies more data to be exchanged, which in turn leads to an increase of sensor nodes' power consumption. Therefore, the less the amount of semantic data to store, the better.

ScenarioB - *Detection of inconsistencies.* This scenario is adapted from the Oil Spill scenario [8] of the European project ENVISION [9]. The project aims at developing an infrastructure for environmental web services with ontologies.

Scenario: accidental oil releases to the sea may have severe consequences on both natural resources and human enterprises. For oil spill decision making, it is essential to be able to predict the fate and effects of the spilled oil. Fate prediction requires data on the spill (location, time, amount, oil type), the environmental conditions (wind, current), and geography (sea depths, coast line). A decision-maker has different datasets for oil spill prediction at his/her disposal and would like to combine them.
Comment: For the purposes of the illustration, it is assumed that - within this scenario - the different types of data for oil spill prediction are available in vector format. It is also assumed that the decision-maker has done a preliminary look-over where all the datasets available were found to have a similar spatial resolution (say 100meters).

Given that there are various ways of defining the spatial resolution for vector data, an additional question to answer is: what criterion is used to assess the spatial resolution of the observation offering? [10] in order to ensure that heterogeneous datasets which have the same data quality value are effectively compatible with respect to their resolution. This check is possible with the ODP proposed through a simple SPARQL query (see details in the Java code).

5 Conclusion

Knowledge about the quality of sensor observations is an important aspect of the discovery of resources in the Semantic Sensor Web. This work has suggested an ontology design pattern (ODP) to characterize the quality of sensor observations. The ODP is relevant for the annotation of sensor observations with spatial data quality components. It can be used to infer spatial data quality component values for existing sensor observations and warn against the integration of sensor observations assessed with different quality criteria. The ODP was aligned to the foundational ontology Dolce Ultra Light and validated through the development of a software program.

[8] See a detailed presentation of the oil spill scenario at `http://envision.brgm-rec.fr/OS-UseCase.aspx` (last accessed: August 30, 2012).
[9] See a presentation of the project at `http://www.envision-project.eu`.
[10] This question is an application of the competency question (ii) from section 3 to the spatial data quality component 'Resolution'.

Acknowledgments

The work is funded by the German Academic Exchange Service (DAAD A/10 /98506), and the European Commission through the ENVISION Project (FP7-249170). It is conducted within the International Research Training Group on Semantic Integration of Geospatial Information (DFG GRK 1498). The ideas presented were shaped through discussions with Patrick Maué, Alejandro Llaves, Werner Kuhn, Alkyoni Baglatzi, Henry Michels and Marcell Roth.

References

1. Barnaghi, P., Meissner, S., Presser, M., Moessner, K.: Sense and sens'ability: Semantic data modelling for sensor networks. In: Cunningham, P., Cunningham, M. (eds.) Proceedings of the ICT-MobileSummit 2009. Santander, Spain (2009)
2. Botts, M., Robin, A.: Bringing the sensor web together. Geosciences 6, 46–53 (2007)
3. Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., Lemmens, R.: New generation sensor web enablement. Sensors 11(3), 2652–2699 (2011)
4. Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W.D., Le Phuoc, D., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K.: The SSN ontology of the W3C semantic sensor network incubator group. Web Semantics: Science, Services and Agents on the World Wide Web (2012)
5. Corcho, O., García-Castro, R.: Five challenges for the semantic sensor web. Semantic Web 1(1), 121–125 (2010)
6. Gangemi, A., Presutti, V.: Ontology design patterns. In: Staab, S., Studer, R. (eds.) Handbook on ontologies, pp. 221–243. Springer Berlin Heidelberg (2009)
7. Goodchild, M.F.: Scale in GIS: an overview. Geomorphology 130(1-2), 5–9 (2011)
8. Hammar, K., Sandkuhl, K.: The state of ontology pattern research: A systematic review of ISWC, ESWC and ASWC 2005-2009. In: Blomqvist, E., Chaudhri, V., Corcho, O., Presutti, V., Sandkuhl, K. (eds.) Proceedings of the 2nd International Workshop on Ontology Patterns - WOP2010. CEUR-WS.org, Shanghai, China (2010)
9. Janowicz, K., Compton, M.: The Stimulus-Sensor-Observation ontology design pattern and its integration into the semantic sensor network ontology. In: Taylor, K., Ayyagari, A., De Roure, D. (eds.) The 3rd International workshop on Semantic Sensor Networks. CEUR-WS.org, Shanghai, China (2010)
10. Kumi-Boateng, B., Yakubu, I.: Assessing the quality of spatial data. European Journal of Scientific Research 43(4), 507–515 (2010)
11. Paradis, J.R., Beard, K.: Visualization of spatial data quality for the decision maker: a data quality filter. URISA Journal 6(2), 25–34 (1994)
12. Sheth, A., Henson, C., Sahoo, S.S.: Semantic sensor web. IEEE Internet Computing 12(4), 78–83 (2008)
13. Veregin, H.: Data quality parameters. In: Longley, P.A., Maguire, D.J., Goodchild, M.F., Rhind, D.W. (eds.) Geographical information systems: principles and technical issues, chap. 12, pp. 177–189. John Wiley and Sons, New York, 2nd edn. (1999)