Provenance in Systems for Situation Awareness in Environmental Monitoring

Markus Stocker, Mauno Rönkkö, and Mikko Kolehmainen

Department of Environmental Science, University of Eastern Finland, Kuopio, Finland
`{markus.stocker,mauno.ronkko,mikko.kolehmainen}@uef.fi`

Abstract. As environmental monitoring systems increasingly automate the collection and processing of environmental sensor network data, the technical components of such systems can automatically obtain and maintain higher levels of situation awareness—awareness of the monitored part of reality. In order to increase confidence in the correctness of situation awareness maintained by such systems it is important to explicitly model provenance. We present an alignment of the PROV ontology with ontologies used in a software framework for situation awareness in environmental monitoring, called Wavellite. The extended vocabulary enables the explicit representation of provenance in Wavellite applications. We demonstrate the implementation for a concrete scenario.

Keywords: Situation awareness, situation theory, provenance, environmental monitoring, Wavellite.

1   Introduction

Endsley defined situation awareness as the "perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [1]. Over decades, situation awareness has been receiving considerable attention in various communities, e.g. human factors and ergonomics [2, 3, 4, 5]. Whether situation awareness is purely an "individual psychological phenomenon" [6] or is distributed between human and technical agents continues to be debated [2, 5, 6, 7]. However, applications of situation awareness theory and systems have been largely limited to military and security domains [2, 8].

Recently, Stocker et al. [9, 10, 11] have adopted Situation Theory [12, 13] and technologies [14] for situation awareness in environmental monitoring. Today, environmental monitoring systems often rely on environmental sensor networks [15] to implement the measurement [16] of properties of physical phenomena over time and space. Data resulting from measurement is collected, processed, and analyzed in order to obtain information about the monitored environment. Environmental monitoring systems are arguably not just technical systems. In fact, on one hand the monitored entities are often organisms or entire ecosystems. On the other hand, people are part of environmental monitoring systems, in roles such as technicians, scientists, or citizens. Thus, environmental monitoring systems may perhaps be described as (enviro-)sociotechnical systems [17].

Situations are structured parts of reality [13]. The concept of situation is interesting in environmental monitoring for at least three reasons. First, the monitored entity is a part of reality. For instance, in urban air pollution monitoring the particular volume of urban ambient air is the monitored entity and is a part of reality. Second, the entity is structured. In our example, the structure of the particular volume of urban ambient air is defined by the relations among the objects that are constituents of the entity, such as particulates and gases. It is properties of such objects that are typically measured. Third, environmental monitoring is concerned with gaining information about the monitored part of reality, i.e. information about situations.

In environmental monitoring, situation awareness is, traditionally, in the mind of people, typically experts. Here, Endsley's model is particularly suitable. However, as the technical components of environmental monitoring systems become increasingly more "intelligent," distributed situation awareness models [5] are of interest as well. Indeed, as the technical components increasingly implement data collection and processing, as well as knowledge extraction and representation, the technical parts can arguably hold higher levels of situation awareness—shared with people, such as scientists.

With increasing automation of data collection, data processing, knowledge extraction, and knowledge representation, it becomes important to automatically model (data) provenance. Provenance enables tracing the processes involved in producing data and knowledge. It can increase confidence in the correctness of situation awareness obtained and maintained by environmental monitoring systems, in particular that of their technical parts. Our aim in this paper is to extend an alignment of ontologies [18] used in a software framework for situation awareness in environmental monitoring, called Wavellite, with the PROV Ontology (PROV-O) [19]. As our main contribution, we present the alignment and discuss its application for a concrete example.

2   Materials and Methods

Wavellite[1] [9, 10, 20] is designed to support the implementation of data collection, data processing, knowledge extraction, and knowledge representation. Collection is, often, for data resulting in measurement implemented by environmental sensor networks. Such data are sensor observations, which are aligned with the term Observation of the Semantic Sensor Network (SSN) ontology [21]. Processing is for dataset observations, which are aligned with the term Observation of the RDF Data Cube Vocabulary (QB) [22]. Knowledge is for situations and is, specifically, situational knowledge. Situational knowledge is extracted from dataset observations and is represented as Situation, a term of the Situation Theory Ontology (STO) [14].

An alignment of these three ontologies, plus OWL-Time [23] and GeoSPARQL [24] for the representation of time and space, respectively, has been proposed in [18]. This alignment forms the Wavellite Core Ontology (WCO) which we extended with Wavellite terms (e.g. SensorObservation) to form the Wavellite Entity Ontology (WEO). Here we modify WCO to include PROV-O, and propose an alignment between PROV-O and

---

[1] http://uef.fi/envi/projects/wavellite

WEO. As a result, PROV-O joins the WCO family of upper ontologies used in Wavellite to represent data, knowledge, metadata, and now provenance. We used Protégé[2] to create the alignment. Note that such alignment is independent of concrete software implementations, such as Wavellite.

PROV-O is a specification for provenance designed for the representation of the origins of digital objects in form of descriptions "of the entities and activities involved in producing and delivering or otherwise influencing a given object" [25]. In PROV-O, provenance is, generally, of entities, which can be physical, digital, or conceptual. Entities can be derived from other entities and they are generated by activities. Activities are the processes through which entities come into existence. Associated with activities are agents, e.g., persons or, of most interest here, software.

In addition to the ontology alignment, we also extend Wavellite such that the software framework supports the representation of provenance in concrete applications. Thus, provenance records can be persisted and retrieved in a similar manner as sensor observations, dataset observations, and situations are persisted and retrieved in Wavellite.

3   Results

In this section, we briefly describe the main elements of the alignment. Sensor observation, dataset observation, and situation are digital objects and, thus, PROV-O entities. Aligning sensor observations with PROV-O is extensively addressed in [26]. The authors propose an alignment that aims at reconciling different aspects of modelling sensor observations in the SSN ontology (constrained by its alignment to the DOLCE Ultralite ontology [27]), OGC Observations and Measurements [28], and PROV-O. As a consequence, the resulting alignment relies on the introduction of several additional classes.

We follow [26] by adopting a lightweight subset of the alignment axioms. Specifically, SSN `Observation` is a sub class of PROV-O `Entity`; SSN `Stimulus` is a sub class of PROV-O `Activity`; and SSN `Sensor` is a sub class of PROV-O `Agent`. Sensor observations are generated by stimuli, are attributed to sensors, and stimuli are associated with sensors. Thus, the SSN object property `observedBy`, used to relate observations and sensors, is a sub property of PROV-O `wasAttributedTo`. In Wavellite, sensor observations are not derived from entities.

Wavellite implements *operators* that translate sensor observations into dataset observations [20]. Translation is an *operation*. Operators are software and thus PROV-O agents. Operations are PROV-O activities. Operations are associated with operators. Dataset observations may be derived from sensor observations, are attributed to operators, and are generated by operations. Wavellite also implements operators that process a source set of dataset observations into a target set of dataset observations. Thus, dataset observations may also be derived from dataset observations. Such operators are associates for the *processing* operation, which uses and generates dataset observations. For instance, the `Aggregate` operator with function `mean` and time period `hour` is a PROV-O agent and associate for the *aggregation* activity that uses

---

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

source sets of dataset observations within one hour window and generates a singleton target set with hourly mean dataset observations. Finally, QB `DataSet` is a PROV-O `Entity` and datasets can thus be derived from datasets.

STO objects, notably situations, elementary infons, relations, individuals, attributes, and values are PROV-O entities. Any of these objects may be derived from dataset observations. In this case, extraction (or acquisition) operations (PROV-O activities) that are associated with extractors, the operators (PROV-O agents), use dataset observations, and generate STO objects. For instance, a classification operation may be associated with a machine learning operator and classify (use) dataset observations to generate (information about) an individual involved in a situations. However, any STO object may also be derived from STO objects. For instance, given information for storms and the location of drivers, a system may infer information for situations in which drivers are at higher risk due to storms [29].

In Wavellite, PROV-O enables the explicit representation of metadata describing the origin of sensor observations, dataset observations, and situations. This is particularly interesting at the derivation layer of the Wavellite architecture, where applications can implement arbitrary complex chains of dataset processing. By modelling datasets and dataset observations as PROV-O entities, we can explicitly model the derivation of datasets and dataset observations from other datasets and dataset observations, respectively, as well as the responsible processes (activities) and involved software (agents). However, provenance is interesting also at the situation layer of the Wavellite architecture, where applications implement the representation of situational knowledge acquired (extracted) from dataset observations. Here provenance enables Wavellite to relate situational knowledge to the dataset observations from which it is derived and with the agents, e.g. data-driven or physically-based models, and activities involved in knowledge acquisition.

4   Discussion

This section discusses a concrete example. The scenario builds on related work [10, 30] and can be summarized as follows. The pavement of a road section is measured for vibration by a sensor network consisting of accelerometers installed into the ground at one side of the road. Occasionally, vehicles travel the road section and modify the vibration pattern measured by sensors. Using a trained artificial neural network, such patterns can be classified in order to detect and characterize vehicles, for instance as 'light' or 'heavy'.

Figure 1 is an example sensor observation in this scenario. As expected, the example relies on the SSN ontology to model (meta-)data about the observation, and on OWL-Time for temporal data. Figure 2 displays the provenance information for the example sensor observation in Figure 1. The sensor observation `ex:44b` is modelled as a PROV-O `Entity` that was generated by the `ex:vibration` PROV-O `Activity` (SSN `Stimulus`) and was attributed to the `ex:sd1` PROV-O `Agent` (SSN `SensingDevice`). The graphs in figures 1 and 2 can be joined via node `ex:44b`.

(a)

(b)
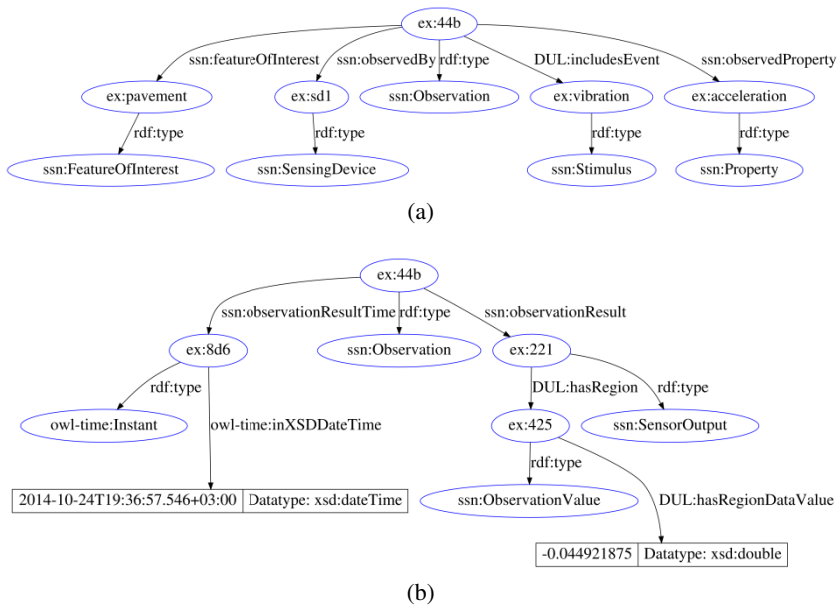
Fig. 1. Sensor observation `ex:44b` that resulted in measurement of road pavement vibration by accelerometer sensing device `ex:sd1` on October 24, 2014 at 19:36:57.546 with observation value -0.044921875. For the sake of readability, the graph is split into two sub graphs. Figure (a) shows the feature, sensor, stimulus, and property related to the observation. Figure (b) shows the temporal location and measurement value. The two sub graphs can be joined via node `ex:44b`.
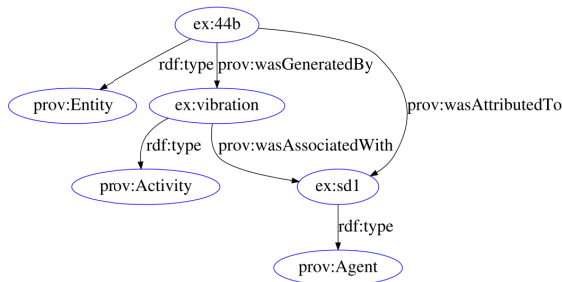
Fig. 2. Provenance information for the example sensor observation (Figure 1). Modelled are in particular the involved PROV-O entity, activity, agent and relations among them.

Sensor observations are translated to dataset observations. Figure 3 shows the result of such translation for our example sensor observation. Dataset observation `ex:3bc` relates to dataset `ex:d1` as well as to time and the acceleration value via two component properties. The graph includes provenance information. It states that the dataset
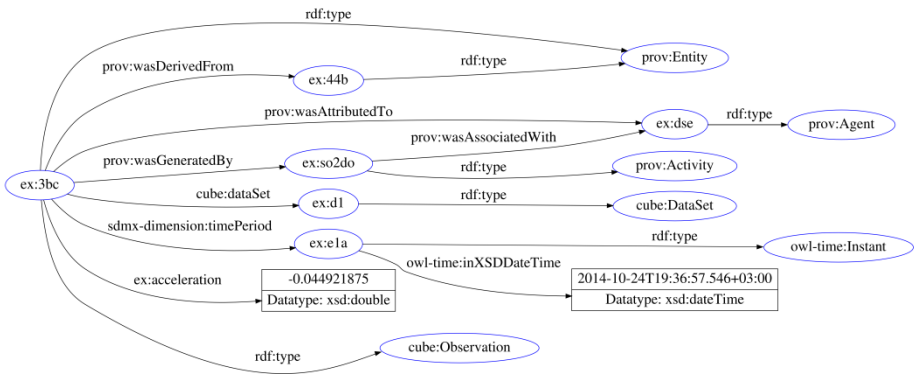
Fig. 3. Dataset observation `ex:3bc` of dataset `ex:d1` with component properties for time and acceleration value. The dataset observation was derived from the sensor observation `ex:44b` and generated in a translation activity associated to a certain agent.

observation was generated by the `ex:so2do` (sensor observation to dataset observation) translation activity associated with the `ex:dse` (dataset engine) agent. The graph also states that the dataset observation `ex:3bc` was derived from the sensor observation `ex:44b`.

In the discussed scenario, for a specified time window length and at regular time intervals, dataset observations in time domain are processed to vibration patterns in frequency domain [10]. Vibration patterns are dataset observations with component property for time and one component property for each represented frequency component. Vibration patterns are then classified using trained Multi-Layer Perceptron artificial neural networks in order to detect and characterize vehicles travelling on the road section. Characterization determines whether the observed vehicle is light or heavy. Detected vehicles are individuals in situations. We can model such situations as supporting an infon with `vehicle-at`-relation and two objects, one for the individual vehicle and the other for the temporal location.

Figure 4 is an example. The situation supports an infon which states that a vehicle (`ex:640`) travelling on the road section was detected at 20:05:36 and was characterized as being light. The example also includes provenance information for the vehicle individual. It states that the individual was derived from the `ex:aa3` PROV-O `Entity` (which an expanded graph would additionally type as dataset observation) and that it was generated in the `ex:classification` PROV-O `Activity` associated with the `ex:se` (situation engine) PROV-O `Agent`.

The discussed example demonstrates how the provenance of information for objects observed in real world situations by an environmental monitoring system can be traced through a complex data processing chain down to the original sensor observations. Provenance also tracks the activities and agents involved in transforming entities.
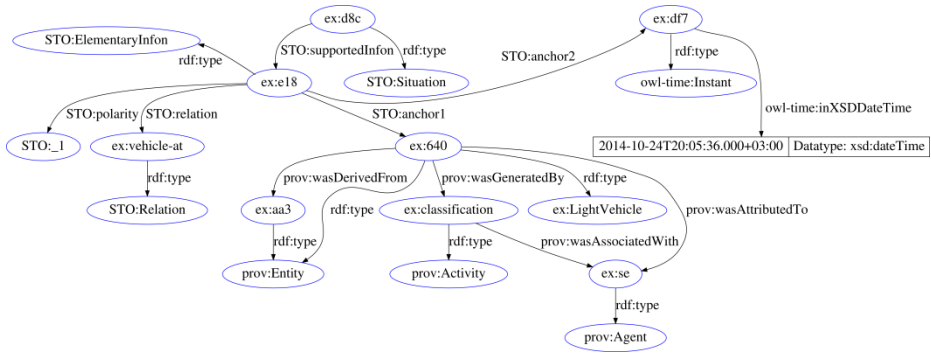
Fig. 4. Situation supporting a `vehicle-at`-relation infon involving two objects, an individual light vehicle and a temporal location. The graph also includes provenance information for the vehicle individual.

Systems can query the RDF data according to whether the interest is for situational knowledge or for the provenance of such knowledge.

5   Conclusion

We presented an alignment of a suite of ontologies useful to situation-aware environmental monitoring systems with PROV-O, the W3C provenance ontology. Related work on the alignment of the SSN ontology and PROV-O proved useful here. We have presented a basic alignment of PROV-O with entities beyond sensor observations, required in situation-aware environmental monitoring systems, namely dataset observations and situations.

The discussed example for situations involving vehicles travelling a road section demonstrates how systems can annotate, during processing, sensor observations, datasets and their observations, and situations with provenance information. Provenance can thus support making transparent the often complex data processing and knowledge extraction chains implemented in situation-aware environmental monitoring systems.

There exist several directions for future work. On one hand, our alignment consists of only few key axioms. More work can be done to study how to improve the alignment. On the other hand, the ideas and implementation presented here can be developed for a concrete system and application. The data and provenance information resulting in such a system can be used to more concretely study the potential of provenance information in a situation-aware environmental monitoring system.

References

1. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society 37(1), 32–64 (1995)
2. Salmon, P.M., Stanton, N.A., Jenkins, D.P., Walker, G.H., Young, M.S., Aujla, A.: What Really Is Going on? Review, Critique and Extension of Situation Awareness Theory. In: Harris, D. (ed.) HCII 2007 and EPCE 2007. LNCS (LNAI), vol. 4562, pp. 407–416. Springer, Heidelberg (2007), `http://dx.doi.org/10.1007/978-3-540-73331-7_45`
3. Smith, K., Hancock, P.: Situation Awareness Is Adaptive, Externally Directed Consciousness. Human Factors: The Journal of the Human Factors and Ergonomics Society 37(1), 137–148 (1995)
4. Bedny, G., Meister, D.: Theory of Activity and Situation Awareness. International Journal of Cognitive Ergonomics 3(1), 63–72 (1999), `http://dx.doi.org/10.1207/s15327566ijce0301_5`
5. Stanton, N.A., Stewart, R., Harris, D., Houghton, R.J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M.S., Linsell, M., Dymott, R., Green, D.: Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics 49(12-13), 1288–1311 (2006), `http://dx.doi.org/10.1080/00140130600612762`, pMID: 17008257
6. Stanton, N.A., Salmon, P.M., Walker, G.H., Jenkins, D.P.: Is situation awareness all in the mind? Theoretical Issues in Ergonomics Science 11(1-2), 29–40 (2010), `http://dx.doi.org/10.1080/14639220903009938`
7. Endsley, M.R.: Automation and Situation Awareness. In: Parasuraman, R., Mouloua, M. (eds.) Automation and human performance: Theory and applications, pp. 163–181. Lawrence Erlbaum Associates, Inc., Mahwah (1996)
8. Salfinger, A., Retschitzegger, W., Schwinger, W.: Maintaining Situation Awareness Over Time – A Survey on the Evolution Support of Situation Awareness Systems. In: Conference on Technologies and Applications of Artificial Intelligence (TAAI 2013), pp. 274–281. IEEE Computer Society (2013)
9. Stocker, M., Baranizadeh, E., Portin, H., Komppula, M., Rönkkö, M., Hamed, A., Virtanen, A., Lehtinen, K., Laaksonen, A., Kolehmainen, M.: Representing situational knowledge acquired from sensor data for atmospheric phenomena. Environmental Modelling & Software 58(0), 27–47 (2014), `http://www.sciencedirect.com/science/article/pii/S1364815214001108`
10. Stocker, M., Rönkkö, M., Kolehmainen, M.: Situational knowledge representation for traffic observed by a pavement vibration sensor network. IEEE Transactions on Intelligent Transportation Systems 15(4), 1441–1450 (2014)
11. Stocker, M., Nikander, J., Huitu, H., Jalli, M., Rönkkö, M., Kolehmainen, M.: Disease pressure situation modelling in agriculture. Computers and Electronics in Agriculture (Submitted, 2014)
12. Barwise, J., Perry, J.: Situations and Attitudes. The Journal of Philosophy 78(11), 668–691 (1981), `http://www.jstor.org/stable/2026578`
13. Devlin, K.: Logic and Information. Cambridge University Press (1991)
14. Kokar, M.M., Matheus, C.J., Baclawski, K.: Ontology-based situation awareness. Inf. Fusion 10(1), 83–98 (2009)
15. Hart, J.K., Martinez, K.: Environmental Sensor Networks: A revolution in the earth system science? Earth-Science Reviews 78(3-4), 177–191 (2006)

16. Finkelstein, L.: Theory and Philosophy of Measurement. In: Sydenham, P.H. (ed.) Handbook of Measurement Science. Theoretical Fundamentals, vol. 1, pp. 1–30. John Wiley & Sons (1982)
17. Walker, G.H., Stanton, N.A., Salmon, P.M., Jenkins, D.P.: A review of sociotechnical systems theory: a classic concept for new command and control paradigms. Theoretical Issues in Ergonomics Science 9(6), 479–499 (2008),
    http://dx.doi.org/10.1080/14639220701635470
18. Stocker, M., Rönkkö, M., Kolehmainen, M.: Towards an Ontology for Situation Assessment in Environmental Monitoring. In: Ames, D.P., Quinn, N.W., Rizzoli, A.E. (eds.) Proceedings of the 7th International Congress on Environmental Modelling and Software, vol. 3, pp. 1281–1288. International Environmental Modelling & Software Society, San Diego (2014), http://www.iemss.org/society/index.php/iemss-2014-proceedings
19. Lebo, T., Sahoo, S., McGuinness, D.: PROV-O: The PROV Ontology. W3C Recommendation, W3C (April 2013), http://www.w3.org/TR/prov-o/
20. Stocker, M., Rönkkö, M., Kolehmainen, M.: A software framework for situation awareness in environmental monitoring. Knowledge-Based Systems (Submitted, 2014)
21. 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., Phuoc, D.L., 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 17(0), 25–32 (2012)
22. Cyganiak, R., Reynolds, D., Tennison, J.: The RDF Data Cube Vocabulary. Recommendation, W3C (January 2014),
    http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/
23. Hobbs, J.R., Pan, F.: Time Ontology in OWL. Working draft, W3C (September 2006),
    http://www.w3.org/TR/owl-time/
24. Perry, M., Herring, J.: OGC GeoSPARQL - A Geographic Query Language for RDF Data. Tech. Rep. OGC 11-052r4, Open Geospatial Consortium (September 2012)
25. Gil, Y., Miles, S.: PROV Model Primer. W3C Working Group Note, W3C (April 2013),
    http://www.w3.org/TR/2013/NOTE-prov-primer-20130430/
26. Compton, M., Corsar, D., Taylor, K.: Sensor Data Provenance: SSNO and PROV-O Together At Last. In: Proceedings of the 7th International Workshop on Semantic Sensor Networks 2014 (SSN2014), 13th International Semantic Web Conference, Riva del Garda, Trentino Italy (October 2014)
27. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Oltramari, R., Schneider, L., Istc-cnr, L.P., Horrocks, I.: WonderWeb Deliverable D17. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology (2002)
28. Geographic information – Observations and measurements (2011),
    http://www.iso.org/iso/catalogue_detail.htm?csnumber=32574
29. Stocker, M., Kauhanen, O., Hiirsalmi, M., Saarela, J., Rossi, P., Rönkkö, M., Hytönen, H., Kotovirta, V., Kolehmainen, M.: A Software System for the Discovery of Situations Involving Drivers in Storms. In: Denzer, R., Argent, R., Schimak, G., Hřebíček, J. (eds.) ISESS 2015. IFIP AICT, vol. 448, pp. 229–237. Springer, Heidelberg (2015)
30. Stocker, M., Rönkkö, M., Kolehmainen, M.: Making Sense of Sensor Data Using Ontology: A Discussion for Road Vehicle Classification. In: Seppelt, R., Voinov, A., Lange, S., Bankamp, D. (eds.) 2012 International Congress on Environmental Modelling and Software, pp. 2387–2394. iEMSs, Leipzig (2012)