

Semantic Model Extensibility in Interoperable IoT Data Marketplaces: Methods, Tools, Automation Aspects

Yulia Svetashova^[0000–0003–1807–107X]

Robert Bosch GmbH, Corporate Sector Research and Advance Engineering
Robert-Bosch-Campus 1, 71272 Renningen, Germany
yulia.svetashova@de.bosch.com
Karlsruhe Institute of Technology, AIFB
Kaiserstr. 89, 76133 Karlsruhe, Germany

Abstract. The proposed research is concerned with the problem of extending semantic models used in the IoT data-sharing contexts with new elements. These suggestions are made by the domain experts at run time and then validated by the ontology engineers. The process of terms suggestion is organized as a modeling dialogue: dynamically generated user interface forms with labels in natural language assist the domain experts in the creation of structured proposals. We hypothesize that this communication procedure will increase the efficiency and the perceived usability of performing change operations at different stages of the ontology evolution process. We then suggest a series of experiments to evaluate our approach.

Keywords: Semantic Interoperability · Internet of Things · Data Marketplace · Ontology Evolution · Ontology-Enhanced User Interface

1 Context and Motivation

The study by International Data Corporation “Data Age 2025” estimates a 10-fold growth of annually created data in 2025; a quarter of this data is predicted to be real-time in nature and 95% of it is expected to originate from the Internet of Things (IoT) [17]. According to Manyika et al. [13], despite this constant increase in the amount of IoT data, the lack of interoperability leads to it being used mostly for anomaly detection and current operations, thereby leaving its prediction and optimization potential untapped.

On the other hand, the data-intensive AI-based algorithms and applications which are intended to transform the everyday lives of people and businesses, have a critical dependency on high quality data which is difficult for them to obtain. Sharing this IoT data can be thus beneficial for both data owners and potential data consumers. To satisfy this increasing demand, a new type of IoT data-sharing platform – the *data marketplace*, has started gaining popularity in recent years [19]¹.

¹ The examples of these marketplaces include: DatabrokerDAO, IOTA data marketplace, AppNexus, Trimble, Datum, MDM (traffic data) and Wibson (personal data).

The IoT data marketplaces trade data streams and/or aggregated sensor data and services. These marketplaces aim at becoming a “central point of discoverability” for IoT data. They are also required to ensure interoperability by defining “metaformats and abstractions” [12] for cross-standard, cross-platform and cross-domain use cases. These “metaformats” (marketplace standards, in our case ontology-based semantic models) serve a **twofold purpose**: 1) *categorization of data offerings* – related or similar offerings are placed closer together and can be retrieved with queries of different levels of generality; 2) *data integration* – input and output parameters of different offerings are expressed in terms of one mediated schema making heterogeneous data jointly usable in the consumer application context.

2 Problem Statement

The creation and use of semantic models for automated IoT resource or service discovery and subsequent data integration at run time presents two main challenges. The first challenge is that these models need to provide unified semantics to capture not only raw sensor data but **context information** [16] which becomes critically important when relevant data is filtered and/or heterogeneous sources are combined. The second challenge is that these models require **constant adaptations** based on the change requests coming from varying users who want to provide data with very different underlying schemata and real world conceptualizations.

An early study of semantic interoperability in decentralized distributed environments by Vetere and Lanzerini [22] stresses that a “suitable coverage of primitive concepts with respect to the business domain (completeness) that entails the possibility to progressively enhance conceptual schemas (extensibility)” is the key success factor in the task of data integration with which we are faced. The majority of current interoperability efforts in IoT reflect this requirement. Nonetheless, these schema enhancement methods of including new model elements, at run time, in the absence of a centralized semantic authority, have not yet been examined in a controlled setting. The current dissertation will address these issues. The results of this study will be the basis of the partially automated algorithm of model extension at run time.

3 Related Work

The present work takes into account the results of current European initiatives in the field of IoT **semantic interoperability**, in particular, the experience accumulated in the BIG IoT² project and semantic modeling solutions tested in the IoT-Lite [2], FIESTA-IoT [1] and Inter-IoT [7] projects. The BIG IoT project follows the approach of reusing existing ontologies by combining them within the data-sharing platform.

² “Bridging the Interoperability Gap in the Internet of Things”, see the description at the project website: <http://big-iot.eu>; the marketplace Web portal: <https://market.big-iot.org>.

The idea of thematic templates (see below in Sec. 6) supporting the model enrichment process extends various concepts of **patterns** circulating in the Semantic Web community: “ontology design patterns” [6], design patterns translated into templates [11], observation-driven “geo-ontology patterns” [10], “ontology alignment design patterns” [18]. The thematic template of a model element is its initial broad categorization and a coherent set of prototypical relations in the enclosing model that characterize elements of this category.

Another relevant research domain for the present study is **ontology evolution** defined as “timely adaptation of an ontology to the arisen changes and the consistent propagation of these changes to dependent artefacts” [20]. Flouris et al. [4] distinguishes 6 ontology evolution sub-tasks: 1) capturing required changes, 2) change representation using a formal language, 3) testing effects of the changes, resolving conflicts and forming a complete change request, 4) change implementation and verification, 5) change propagation to dependent data and affected applications, and finally, 6) change validation by the ontology engineer. Current research is planned to contribute mostly to the sub-tasks (1) and (2), but the effects of our approach on other evolution phases will be also investigated.

The proposed ontology-enhanced user interface [15] reflects the conception of the **modeling process “as a dialogue** between participants”, “rooted in the sharing, translation, negotiation, argumentation, and imposing of (participant-based) knowledge representations” [9]³. We adapt this approach to the new concept suggestion procedure.

The research on **distributed collaborative tagging** systems and, in particular, the investigation of the underlying cognitive mechanisms formalized in the semantic imitation model of social tag choices by Fu et al. [5] has an additional influence on our work.

4 Research Questions

We understand the general approach to Semantic Web technologies-based information sharing to be “attaching semantic meta-data to information items, and (...) relating these metadata to each other through background knowledge in the form of ontologies” [21]. On this basis, this study will address the following research questions related specifically to the model enrichment process:

- **RQ1** How can we simplify the process of semantic model extension at run time for domain experts who propose new model elements⁴?
- **RQ2** How can we simplify the process of semantic model extension for ontology engineers who validate new element proposals created by the domain experts at run-time?
- **RQ3** Which aspects of the new element inclusion and validation processes can be automated and to which degree?

³ See also an approach to knowledge authoring as dialogue in Parvizi et al. [14].

⁴ Particularly, we concentrate on those domain experts who have only limited exposure to the model and who are not knowledge modeling experts.

5 Hypotheses

The hypotheses corresponding to our research questions are the following:

H1: The organization of the semantic model extension process as a modeling dialog (coupled with data annotation process) by using an ontology-enhanced user interface where modeling primitives (basic concepts), related model fragments and similar offerings are made accessible, can simplify the process of suggesting the new concepts by domain experts.

H2: The resulting structured proposals, where concepts are linked to the corresponding super-classes and relations are established with other relevant concepts, can simplify the process of the validation and approval of new elements by ontology engineers.

H3: The proposal validation process can be, at least partially, automated by a machine learning classifier. Validation can be modeled as a single-label multiclass classification task where a set of categories {“Accept”, “UseExisting”, “ModifyAndAccept”} represent the decisions⁵. The features for model training are derived from the model element characteristics (structured annotations) proposed by the domain expert, the context of a suggestion and the computed statistics of all marketplace offerings.

6 Approach and Preliminary Results

Empirical and conceptual findings of the proposed research as well as the resulting software artifacts will be an extension of the BIG IoT project. We first outline the relevant model design and system architecture considerations of the project and then indicate the novel components we propose.

There are three groups of the semantic modeling artefacts in BIG IoT. The Core ontology represents the concepts relevant for the marketplace operation: *core:OfferingQuery*, *core:Provider*, *core:Consumer*, *core:Offering*, etc.). Two domain ontologies – Mobility and Environment – provide terms to characterize the content of the data/service offering: *mobility:ParkingSiteStatus*, *environment:PollutionIndicator*, *mobility:TrafficSpeed*, etc.). Finally, the Application ontology is used for organizing navigation through the model elements in the marketplace Web portal⁶.

The marketplace Web portal is used to assist data providers in describing their data and data consumers in formulating queries (see Fig. 1). The user starts by selecting a category and a subcategory, and based on this initial choice, the set of available annotations for the output data is shown in dropdown lists. If users do not find a matching semantic term, they can propose a new one by typing the concept name.

⁵ I owe the idea of modeling the subject expert decision as a classification task to Dmitry Mironov (private conversation).

⁶ The models follow Linked data principles and are published as schema.org custom extensions; the example terms are dereferenceable provided that the namespace prefixes can be expanded as follows: *core* – <http://schema.big-iot.org/core/>, *mobility* – <http://schema.big-iot.org/mobility/>, *environment* – <http://schema.big-iot.org/environment/>. Other prefixes in the paper are used as declared on prefix.cc

The new term is assigned a namespace prefix “proposed” and stored as a metadata annotation for the current offering and as a potential model element. This will be later approved by an ontology engineer responsible for the marketplace models maintenance.

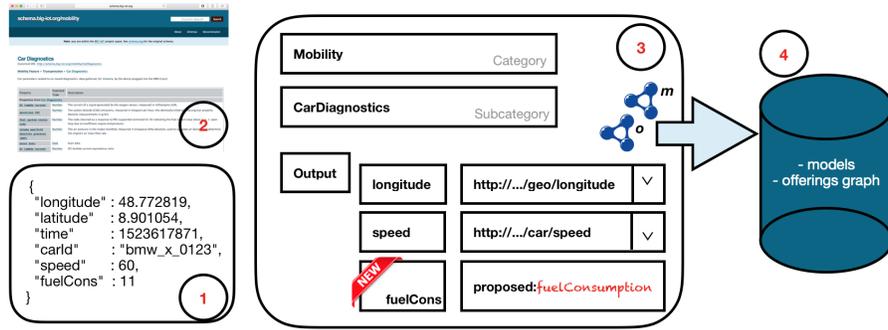


Fig. 1. Schematic representation of the data description process (current state). Given the (1) output sample to be described as a data offering and (2) the ontology model to be used, the marketplace user interface (3) provides a set of expected annotations depending on the selected category. The resulting description in RDF format is stored in the triple store (4) as a part of the offerings metadata graph.

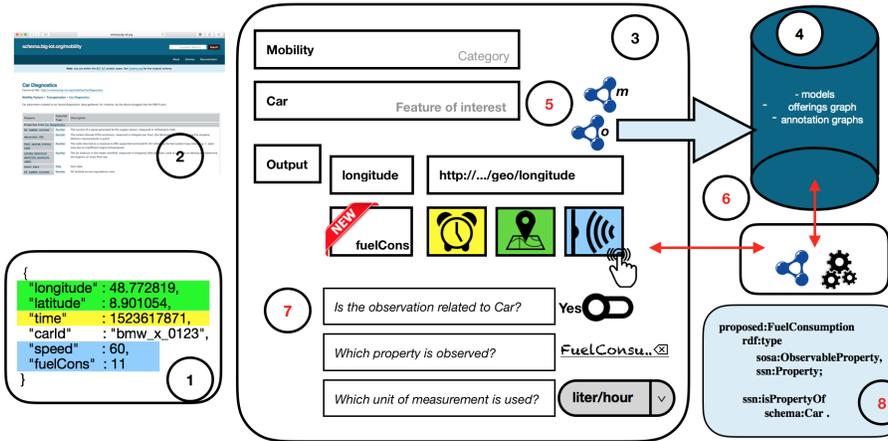


Fig. 2. Schematic representation of the proposed data description and model extension process. Components (5) – (8) marked red, point to the main modifications in comparison with the current state of the system.

Addressing **RQ1**, we suggest the following modifications of the new model element inclusion process (they are summarized in the Fig. 2 above):

(1) All marketplace domain ontologies are aligned with the SSN/SOSA ontologies [8]; the **categorization** of the offerings is coupled with the concept of *sosa:FeatureOfInterest* (⑤).

(2) Data-driven **thematic templates** (we start with temporal, spatial and sensor measurement templates), are derived from a representative set of data, matching the marketplace scope, and stored as a collection of annotation graphs in a triple store (the colored layering in the sample output ① in Fig. 2 reflects thematically different data).

(3) Proposing a new model element, the user first selects a thematic template which provides a high level characterization of the proposed concept. The system, in turn, sends a request to the back-end module ⑥ whose main function is to retrieve a corresponding **annotation graph** from a triple store and to dynamically generate a series of **web forms** ⑦ based on the structure of the annotation graph (the process is exemplified in Table 1).

Table 1. Modeling as dialogue: triples from the sensor measurement annotation graph are transformed into questions in natural language which become labels of the corresponding user interface elements. Basic graph pattern with variables in Turtle syntax.

Annotation Triple(s)	UI Label	UI Element
sosa:Observation sosa:hasFeatureOfInterest ?FeatureOfInterest .	“Is the observation related to <i>Car</i> ?” (actual <i>sosa:FeatureOfInterest</i> selected)	Toggle switch
sosa:Observation sosa:observedProperty ?ObservableProperty .	“Which <i>property</i> of <i>Car</i> is observed?”	Text field
sosa:Observation sosa:hasResult sosa:Result [rdf:type qudt:QuantityValue [qudt:unit ?Unit]].	“Which <i>unit of measurement</i> is used?”	Dropdown list with units of measurements.

(4) By answering questions displayed in the web forms, the user creates a **structured semantic annotation** of the new concept ⑧. The annotation is persisted in the triple store as a new model element with the namespace prefix “proposed”.

To test **H2** corresponding to the **RQ2**, we will build a user interface for the ontology engineers responsible for the maintenance of the marketplace models. This interface will allow interactive exploration of 1) structured user proposals; 2) their contexts (the offering descriptions), 3) thematic templates and existing descriptions based on them.

In our setting, we are interested in incorporating complex user-initiated changes. These changes involve the addition of new model elements, the es-

establishing of an is-a relation with existing concepts as well as the introduction of the non-taxonomic property relations specified in the user annotation.

An ontology engineer may: 1) accept the proposed concept and the annotation without any modification, 2) decline a change request and use the existing model element instead, or 3) accept a concept but change the proposed annotation. In the cases when a change is accepted (1 or 3), it becomes a part of the model, and the effects of this change are propagated to all dependent/related offerings and taken into account by the matching queries. When the proposed concept is declined (2), the offering metadata is corrected accordingly.

Our preliminary results in the BIG IoT project include the alignment of the domain models with the SSN/SOSA framework, the data-driven investigation of archetypal thematic templates and the translation of their prototypical representation into basic annotation graphs. The first version of the system architecture extension (the UI dynamic form generation and the corresponding back-end) is implemented.

7 Evaluation Plan

The proposed approach and associated hypotheses will be evaluated in a series of experiments with the groups of users representing the two roles in the change management procedure: the *domain experts* group proposing changes and the *ontology engineers* group validating the proposed changes.

We have started by the construction of a *gold standard* dataset: 60 data samples in JSON format were annotated by two ontology engineers with a category and a data type for each key-value pair in the sample. Next, a selection of samples will be formed based on the following criteria: the data samples should be 1) typical for the intended marketplace scope; 2) balanced in terms of domains and *sosa:FeaturesOfInterest*; 3) represent various thematic templates.

Then we will intentionally exclude a subset of elements representing *hapax legomena* (i.e. terms which occur only once in our gold standard corpus) from the existing domain models. The task formulated for a group of domain experts is to annotate given samples using the published domain models in the BIG IoT marketplace Web portal. If an annotation term is not found in the available domain models, a user is supposed to extend the model.

Experiment 1 is designed to test **H1**, namely, to explore a causal relationship between the ease of suggesting a new concept for data providers (dependent variable) and their exposure to the “modeling as dialogue” approach (causal variable). The control group will use the BIG IoT marketplace Web portal described in Fig. 1. The experimental group will use the implementation of the proposed approach illustrated in Fig. 2. Then the groups will be compared in terms of 1) time spent on performing an annotation task (overall and per sample) and 2) the number of times the model websites are consulted (overall and per sample). Additionally we will look at 3) the agreement with the gold standard annotation for the terms present in the model; 4) the inter-rater agreement within the groups; 5) the naming strategies used for labeling the new concepts.

We will also collect reflections of the group members on how the usability of proposing a new concept is perceived, and compare them in a structured way. For the experimental group, the average agreement with the gold standard annotation for the new terms will be quantified.

Experiment 2 is designed to test **H2**, namely, to prove a causal relationship between the ease of validating a new concept proposal by the ontology engineer (dependent variable) and the prior usage of the “modeling as dialogue” approach by domain experts (causal variable). Ontology engineers will validate proposals from the control group and structured proposals from the experimental group. The validation process will be compared for 2 groups of proposals in terms of 1) time spent on performing a validation task (overall and per sample), 2) the steps needed to integrate a proposed change to the corresponding domain model (per change), 3) the steps needed for change propagation (per change); and 4) the inter-rater agreement between ontology engineers. Questionnaires based on the System Usability Scale [3] will also be used to structure the perceived difficulty/ease of integrating changes, and specifically when dealing with structured proposals.

Based on the validation results with high inter-rater agreement, a new dataset that models the decision making task will be compiled to test **H3**. In a **machine learning experiment** which aims at predicting an ontology engineer decision (dependent variable) based on the features sketched in the previous section, the standard evaluation measures of precision, recall and their harmonic mean F1 score will be used to test the model performance on a test subset.

8 Reflections

The proposed research introduces the problem of run-time semantic model extension by a domain expert mediated by the ontology-enhanced user interface and further validated by an ontology engineer. Its main contribution is intended to be the experimental and theoretical basis for extending semantic models. These models will serve both as descriptive metadata for the resources exposed on an IoT data marketplace and also as a lingua franca (shared model) in the infrastructure where data sharing takes place.

Through the exploration of various scenarios of on-the-fly model extension performed by data providers, we will develop methods and tools to optimize the new model element inclusion process, paying special attention to the aspects which can contribute to the partially automated workflow. Although our research is shaped by and tailored to the IoT data marketplaces context, we believe that the approach described above is generic and can be applied in many scenarios that require model usage and extension by non-experts.

Acknowledgements. This work has been developed in the project BIG IoT funded by the European Commission’s Horizon 2020 program under the grant agreement No 688038. The author would like to thank Prof. Dr. York Sure-Vetter (the thesis advisor), Prof. Dr. Andreas Harth, Dr. Stefan Schmid and three anonymous reviewers for their critical reading of the paper, their constructive comments and suggestions.

References

1. Agarwal, R. et al.: Unified IoT Ontology to Enable Interoperability and Federation of Testbeds. *IEEE 3rd World Forum on Internet of Things (WF-IoT)* (2016): 70–75.
2. Bermudez-Edo, M. et al. IoT-Lite: a lightweight semantic model for the Internet of Things. *Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress, IEEE* (2016): 90–97.
3. Brooke, J.: SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4–7 (1996).
4. Flouris, G. et al.: Ontology Change: Classification and Survey. *Knowledge Engineering Review*, **23**, 117–152 (2008).
5. Fu, W. T. et al.: A semantic imitation model of social tag choices. In: *Computational Science and Engineering. CSE'09. International Conference on*. Vol. 4, 66–73, (2009).
6. Gangemi, A., Presutti, V.: Ontology design patterns. In: *Handbook on ontologies*, 221–243. Springer, Berlin, Heidelberg (2009).
7. Ganzha, M. et al.: Semantic interoperability in the Internet of Things: an overview from the INTER-IoT perspective. *Journal of Network and Computer Applications*, **81**, 111–124 (2017).
8. Haller, A. et al.: Semantic Sensor Network Ontology. W3C Recommendation. W3C. <https://www.w3.org/TR/vocab-ssn/>. Accessed: April 11 2018.
9. Hoppenbrouwers, S. et al.: A fundamental view on the process of conceptual modeling. In: *International Conference on Conceptual Modeling*, 128–143 (2005).
10. Janowicz, K.: Observation-driven geo-ontology engineering. *Transactions in GIS* **16**(3), 351–374 (2012).
11. Jupp, S. et al.: Populous: a tool for building OWL ontologies from templates. *BMC bioinformatics* **13**(1) (2012).
12. Deichmann, J. et al.: Creating a successful Internet of Things data marketplace. (2016). <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/creating-a-successful-internet-of-things-data-marketplace>. Accessed: April 11 2018.
13. Manyika, J. et al.: *The Internet of Things: Mapping the value beyond the hype*. McKinsey Global Institute. (2015).
14. Parvizi, A. et al.: A pilot experiment in knowledge authoring as dialogue. *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Short Papers* (2013).
15. Paulheim, H., Probst, F.: Ontology-enhanced user interfaces: A survey. *Semantic-Enabled Advancements on the Web: Applications Across Industries: Applications Across Industries*, **214** (2012).
16. Perera, Ch. et al.: Context-aware computing for the Internet of Things: A survey. *IEEE communications surveys & tutorials* **16**(1), 414–454 (2014).
17. Reinsel, D. et al.: Data age 2025: The evolution of data to life-critical. (2017). <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>. Accessed: April 11 2018.
18. Scharffe, F. et al.: Ontology alignment design patterns. *Knowledge and information systems*, 40(1), 1–28 (2014).
19. Stahl, F. et al.: A classification framework for data marketplaces. *Vietnam Journal of Computer Science* **3** (3), 137–143 (2016).
20. Stojanovic, L.: *Methods and Tools for Ontology Evolution*. PhD Thesis, University of Karlsruhe, Germany, 2004.

21. Stuckenschmidt, H., Van Harmelen, F.: Information sharing on the semantic web. Springer Science & Business Media (2005).
22. Vetere, G., Lenzerini, M.: Models for semantic interoperability in service-oriented architectures. *IBM Systems Journal* **44** (4), 887–903 (2005).