District-Scale Data Integration by Leveraging Semantic Web Technologies: A Case in Smart Cities

Kiril Tonev, Simon Kappe(✉), Preslava Krahtova, Hendro Wicaksono, and Jivka Ovtcharova

Karlsruhe Institute of Technology, Institute for Information Management in Engineering, Zirkel 2, Karlsruhe, Germany
{kiril.tonev,simon.kappe,preslava.krahtova,hendro.wicaksono, jivka.ovtcharova}@kit.edu
http://www.imi.kit.edu

Abstract. Technologies of the Semantic Web stack promise to alleviate some of the challenges related to data integration on a massive scale and high level of heterogeneity. This paper explores their application in the smart cities domain with a focus on energy efficient districts. We develop an ontology grounded in several well-established vocabularies to leverage their shared semantics and facilitate data interoperability and we apply the developed ontology to integrate state-of-the-art energy simulation facilities into a general district-level monitoring framework.

Keywords: Smart cities · Building energy simulation
Semantic data integration · Ontology alignment

1 Introduction

The energy management at a city or district level requires a synergetic participation of multiple stakeholders, for instance citizens, utility companies, policy makers or municipalities, and companies providing services related to energy efficiency e.g. energy consultants, building energy management providers, software companies. Multiple tools or services are also utilized to help the stakeholders to perform energy management activities. The multiple stakeholders and services consume and produce data resulting in heterogeneous data formats and sources. The data often come from multiple and unrelated domains, for example building geometry and topology data, sensor data, occupant behavior data, regulations, weather and geospatial data [1].

In order to accomplish accurate and efficient energy management activities involving multiple stakeholders, it requires access and efficient processing of up-to-date and comprehensive information coming from the heterogeneous data [2]. The interoperability issue is still the main challenge in achieving it. Semantic

The original version of this chapter was revised: This chapter was originally positioned in the back matter as a short paper. It has been moved to the main part of the publication. The correction to this chapter is available at https://doi.org/10.1007/978-3-319-73805-5_35

C. Debruyne et al. (Eds.): OTM 2017 Workshops, LNCS 10697, pp. 289–292, 2018.
https://doi.org/10.1007/978-3-319-73805-5_31

Web technologies promise to alleviate some of the challenges related to data integration on a massive scale and high level of heterogeneity. This paper introduces an approach to develop an ontology that reuses well-established vocabularies to leverage their shared semantics and facilitate data interoperability in the smart city context.

The energy consumption analysis on district or city level requires an aggregation of energy consumption data of buildings located in the district. Since not all buildings are equipped with energy metering infrastructure, building energy simulations are performed to predict the energy consumption of a building based on building envelope profile, occupant profile, and surrounding factors.

2 Ontology Development

In DAREED project, we have followed the principle of ontology reuse and grounded the main concepts within the frameworks of other, well-established ontologies. Figure 1 summarizes the main concept of the ontology.

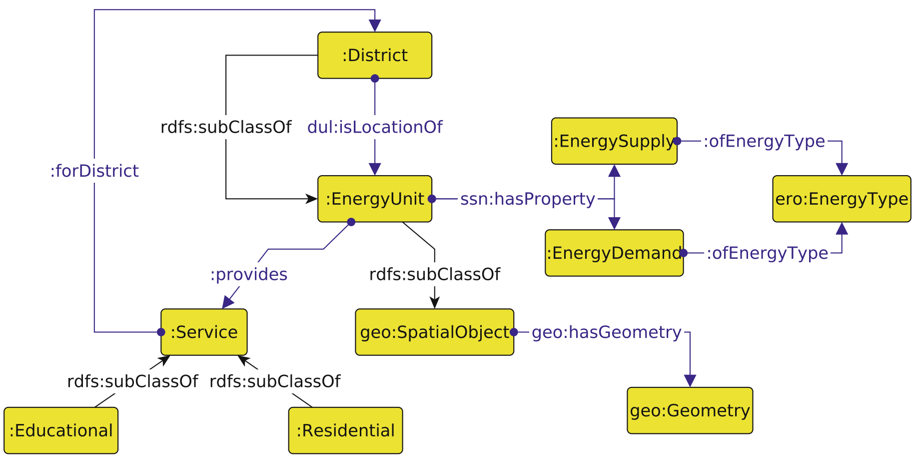

Fig. 1. Main concepts of DAREED district ontology.

In our model, a district is described by a region encompassing a set of energy units. The demand and supply of the district are dependent on the energetic behavior of the energy units within. To further distinguish energy units, we describe the main services they fulfil as members of the district: for example, houses provide residential services, schools provide educational services, parks provide open spaces, and office buildings provide industrial services and so on. An energy unit can demand and supply energy, and we distinguish the types of energy involved in the balance.

The computational platform uses the aggregated area usages for each service category and estimates the demand and response curves of the district in 15-min steps. Since the spatial characteristics of the energy units have such a central role within the district model, we have opted to model these dimensions against the GeoSPARQL ontologys Spatial Object and Geometry concepts.

Other aspects of the upper district ontology are the Key Performance Indicators and metrics that describe the services and districts themselves. To that end, we have followed the suggestions outlined by Fox [3]. We derive the service and district metrics from the Quantity and Measurement concepts of the Ontology of units of measure [4].

3 Integrating Simulation Data Within the District Knowledge Graph

Smart meters are becoming increasingly available in districts as electrical networks are moving towards implementing smart grids for meeting the environmental challenges. Despite that, monitoring a significant portion of the districts energy usage with sensors or smart metering devices remains unviable. We maintain that without addressing the privacy issues, the district-wide coverage of the energy consumption will unlikely penetrate the household segment.

To cope with that challenge, we have developed a framework for integrating simulation data directly into the district knowledge graph. Thereby, we have leveraged EnergyPlus simulation package since its availability as open source and text-based IO formats.

Our approach bases on [5], in that it leverages a mapping specification for providing a linked-data view of the simulation results. The mapping specification is XML-based and offers two distinct operators not available in existing packages: the aggregation operator building a single value from a range of fields and the list operator exporting a range of fields as `rdfs:List`.

The mapping specification facilitates the extraction of RDF graphs from table-based sources. It proposes a way for specifying the URLs of resources and their property assignments in an XML file. Once a specification is written, a graph can be extracted by processing the input schemas or tables with an implementation of two interfaces which instruct how to access the data items and their properties.

Given an input and an output file, we proceed to extract a linked-data graph describing the energy unit and its energetic behavior. The mapping file we developed for the purposes of the DAREED project can be logically subdivided into three categories: building structure (or product model), the resources model and the sensing model.

The resources model describes the energy producers and consumers represented in the simulation input/output data. From the EnergyPlus data dictionary, we extract Simulation Resources entities expressing different level of aggregation of the calculated energetic behavior. These range from the Zone to the Plant level. We extend the Semantic Sensor Network Ontology [6] with the special EnergySupply and EnergyDemand concepts expressing which side of the energetic balance the calculations should be related to. Lastly, we further describe the type of energy involved in the energetic balance of the building by drawing from the Energy Resources Ontology.

With the resources model, we are able to describe the energetic behavior of the simulated building or individual zones thereof only qualitatively. The actual estimated measurements are the subject of the Sensing Model. The Sensing Model leverages several ontologies and linked data vocabularies to provide a tabular view of the data encoded within the EnergyPlus simulation output without sacrificing the rich metadata model we have described thus far.

This concludes our overview of the conceptualization of a simulated building and the related output data as used in the DAREED project. It allows us to answer the key question required for the functioning of the monitoring platform, namely how the consumption and production of the building and its zones are behaving over time. By selecting appropriate ontologies and vocabularies from semantic web repositories and the literature, we are capable of building a semantically rich model that can be interpreted and processed by third-party systems with very little effort. We emphasize that our model has been primarily informed by the requirements of the project and, more importantly, the simulation input files developed therein.

Acknowledgments. The DAREED project has received funding from the Seventh framework programme of the European Union (EU) under grant agreement 609082.

References

1. McGlinn, K., et al.: Identifying use cases and data requirements for BIM based energy management processes. In: CIBSE Technical Symposium (2016)
2. Wicaksono, H., Dobreva, P., Häfner, P., Rogalski, S.: Methodology to develop ontological building information model for energy management system in building operational phase. In: Fred, A., Dietz, J.L.G., Liu, K., Filipe, J. (eds.) IC3K 2013. CCIS, vol. 454, pp. 168–181. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46549-3_11
3. Fox, M.S.: City data: big, open and linked. In: Municipal Interfaces, pp. 19–25 (2013)
4. Rijgersberg, H., van Assem, M., Top, J.: Ontology of units of measure and related concepts. Semant. Web **4**(1), 3–13 (2013)
5. Skjveland, M.G., Giese, M., Hovland, D., Lian, E.H., Waaler, A.: Engineering ontology-based access to real-world data sources. Web Semant. Sci. Serv. Agents World Wide Web **33**, 112–140 (2015)
6. Compton, M., et al.: The SSN ontology of the W3C semantic sensor network incubator group. Web Semant. Sci. Serv. Agents World Wide Web **17**, 25–32 (2012)