Towards a Solution for an Energy Knowledge Graph

Dušan Popadić[1,2], Enrique Iglesias[3][0000-0002-8734-3123], Ahmad Sakor[3][0000-0001-8007-7021], Valentina Janev[2][0000-0002-9794-8505], Maria-Esther Vidal[3,4] [0000-0003-1160-8727]

[1] School of Electrical Engineering, University of Belgrade, Serbia
[2] Mihajlo Pupin Institute, University of Belgrade, Serbia
[3] L3S Research Centre, Leibniz University of Hannover, Germany
[4] TIB-Leibniz Information for Centre for Science and Technology, Germany
dusan.popadic@pupin.rs, iglesias@l3s.de, valentina.janev@pupin.rs, vidal@l3s.de

Abstract.
The recently adopted EU strategy for energy system integration calls for a more integrated energy infrastructure based on innovative technologies. The modernization of the energy sector also aims to solve the problem of the current fragmented applications (built against energy data silos) and enable data sharing within energy communities by leveraging energy data spaces and semantic technologies as a crucial technology for interoperability. This paper addresses the challenges of energy data management and discusses the process of creating a knowledge graph, for example, energy data space, motivated by the needs of the stakeholders from Serbia and related to the integration of a large number of different renewable energy sources (RES) with the proprietary SCADA system of the Institute Mihajlo Pupin. The Energy Knowledge Graph (KG) has been built by reusing the energy-based semantic data model and the SDM-RDFizer, an open-source tool and interpreter of the W3C Recommendations Standard R2RML and its RDF Mapping Language (RML) extension. The Energy KG has been deployed on a Smart Grid Architecture Model (SGAM) – compliant platform hosted at the Institute Mihajlo Pupin.

Keywords: Energy; Knowledge Graph; Mapping rules; Application; Services.

1. Introduction

Knowledge Graphs (KGs) are key trends among the next wave of technologies [1]. It consists of a collection of interrelated descriptions of entities, objects, events, or concepts (relevant semantic metadata) and a framework for data integration, unification, analysis, and sharing [2]. The technology has been more widely applied in practice after the announcement of Google Knowledge Graph [3] in 2012. Many companies started to explore the technology to gain competitive advantages, mainly in the integration of distributed resources over the Internet. Lately, enterprise knowledge graphs are also used to depict a solution to a concrete real business problem, e.g., for facilitating product/service discovery [4], emergency management [5], or managing knowledge in the energy sector [6].

1.1 The European Electricity System

In the European Union (EU), electrical systems have experienced significant changes driven by EU standard rules for the electricity internal market and climate action plans. The increasing generation of data from distributed renewable data sources creates data integration and processing challenges. Therefore, there is a need to develop computational methods for ingesting, managing, and analyzing big data. More importantly, considering the bidirectional flow of information and energy in Smart Grids, knowledge needs to be extracted from this data, to uncover actionable insights. Hence, the future energy infrastructure will be based on intelligent power electronics, smart meters, context-aware devices, IoT, and AI-driven services. Interoperability problems caused by currently fragmented applications will be overcome in the new generation of grids, thus, enabling data exchange between different players in the energy sector. For instance, the EU data strategy envisages the establishment of energy data spaces based on semantic web technologies and W3C standards. The information model proposed in the context of the International Data Space includes exemplary data models for describing datasets and services metadata needed to facilitate information search, service matching, and data exchange.

1.2 Example Case Study

The recently adopted EU energy-related strategies create opportunities to modernize the energy system, making it competitive and environmentally sustainable. Herein, we will use the example of the Serbian electricity system (see Figure 1).

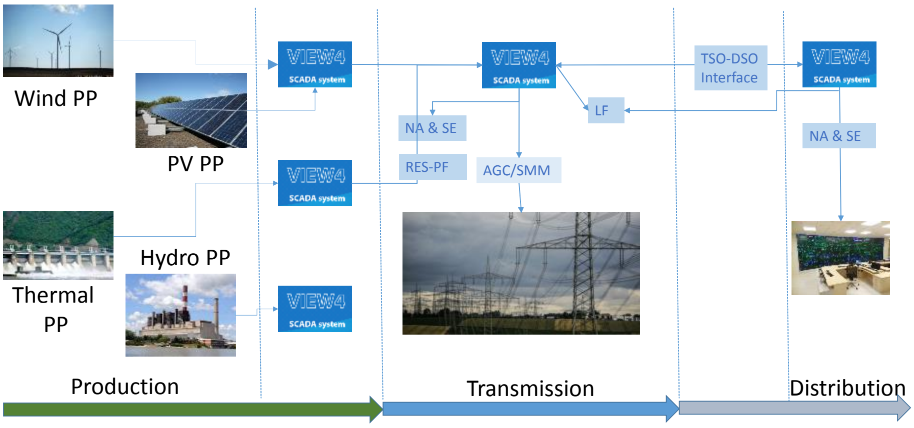

Fig. 1. The Serbian electricity system and the deployment of the VIEW4 SCADA system.

Because the national electricity infrastructure is not isolated, interoperability should be ensured at different levels (i.e., legislation, functional, syntactic, and semantic) and in different parts of the energy value chain, i.e., electricity generation, transmission and consumption. The Institute Mihajlo Pupin proprietary SCADA system has been deployed at many parts of the national electricity grid. The system monitors and controls energy production, distribution, and usage with different objectives, including improving energy efficiency, increasing flexibility and renewable generation share and reducing energy costs. Therefore, the goal of the case study is to provide an innovative energy management service layer on top of existing SCADA based on reusable semantic models or knowledge graphs. They will facilitate the integration of data silos and their fine-grain semantic description; further, they will provide a common understanding of the energy domain based on existing vocabularies.

This paper comprises four additional sections. Section 2 presents the methodology followed to design the pipeline for knowledge graph creation depicted in Section 3. The knowledge graph exploration tools are described in Section 4. Finally, conclusions and future work are outlined in Section 5.

2. Achieving Semantic Interoperability

Interoperability and the possibility of building cross-border and cross-sector services are the focus of many initiatives in Europe; see, for instance, ISA^2 [7]. The high-level vision of the European Union for 2030 is to create a single internal market through a standardised laws system transposed in the national legislation of all member states and a single European data [6] space for data exchange. In order to drive data-based innovations, standardization [8] should be applied, for instance, using metadata schemata, data representation formats, license terms for data and services, data integration [9] and data exchange approaches.

2.1 Research Questions

The Institute Mihajlo Pupin (PUPIN) currently hosts several SGAM (Smart Grid Architecture Model [10]) compliant service-oriented, cloud-based platforms that serve for testing different energy-specific scenarios, see also [11]. Data exchange with external components (e.g., edge computers) is based on an adaptable gateway built upon OGEMA (Open Gateway Energy Management) framework. The data exchange within the broader EU energy ecosystem is still under elaboration. For instance, in the PLATOON project framework, the platform shall be integrated with the PLATOON marketplace based on the Industrial Data Space concept, i.e., using the IDS information model and Linked Data principles [12]. The Semantic Web community has developed more than 700+semantic vocabularies (see the LOV repository, https://lov.linkeddata.es). The aim is to analyze the standard schemas (i.e., vocabularies/ontologies) promoted by the community and adopt them for the targeted SCADA Knowledge Graph and services/applications. The following research questions guide our research:

- RQ1 – Which are the concepts and properties that characterize the energy domain, and which ontologies cover the needs for modelling the electricity value chain and ensure uniform access to data collected with the proprietary SCADA system?
- RQ2 – How to build a knowledge graph that will enable the development of services to support future energy marketplaces?

2.2 Selection of Semantic Models

In order to implement the “no-vendor lock-in” principle and ensure that future services will integrate smoothly with different legacy and proprietary solutions [13], the knowledge graph layer shall be based on open standards and open APIs. Therefore, in our research, one of the first steps toward developing the knowledge graph is the analysis of existing semantic models already in use, such as CIM[1], SAREF[2], SEAS [14] and DCAT[3], defined as following:

- CIM - Common Information Model (CIM), a standard developed by the electric power industry that has been officially adopted by the International Electro technical Commission (IEC); it comprises concepts (e.g., classes or relations) for software applications to exchange information about electrical networks.
- SAREF - Smart Appliances REFerence ontology (SAREF). It is modular ontology for the internet of things domain; it integrates a family of vocabularies to represent smart cities, buildings, energy, agriculture, food, and environmental. SAREF4ENER is an extension for the energy domain; it includes majority of classes of interest for smart energy management.
- SEAS - Ontology developed in the framework of the Smart Energy-Aware Systems (SEAS, https://w3id.org/seas/) project with the aim of designing a global ecosystem of services and smart things collectively capable of ensuring the stability and the energy efficiency of future energy grids. SEAS includes features of interest and their properties, evaluation of features, smart and microgrids, smart homes, electrical cars, electrical market, and weather forecast.
- DCAT - The Data Catalog Vocabulary (DCAT) provides a common understanding of the classes and properties that describe a catalog of datasets and data services. DCAT is expressed in RDF and provides unified representation of catalog properties in a way that is understandable by humans, and also by machines. DCAT includes also classes from other vocabularies, e.g., *foaf:Agent*, *skos:Concept*, or *skos:ConceptSchema*.

2.3 Methodology

The work has been divided into the following phases:

- *Requirement Analysis* phase: the authors, defined different business questions that we would like to answer with the knowledge graph;
- *Design* phase: relevant concepts are selected for modeling. Then, data connectors towards the SCADA database and the messaging mechanisms are specified.
- *Specification* phase: the knowledge graph is specified in terms of RML rules.
- *KGs in Action* phase: the authors are involved in automating the semantic pipe-

---

[1] https://ontology.tno.nl/IEC_CIM/

[2] https://saref.etsi.org/saref4ener/v1.1.2/

[3] https://www.w3.org/TR/vocab-dcat-2

line and developing exploration GUIs.

In this paper, the authors focus on the last two activities, namely the implementation of the semantic pipeline and the exploration of the knowledge graph.

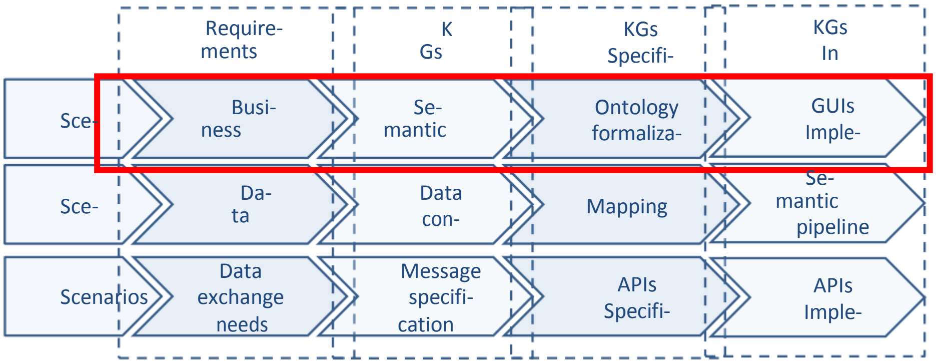

Fig. 2. Four-step Methodology.

3. Knowledge Graph Creation – the Semantic Pipeline

This section describes the process of knowledge graph creation and highlights the main challenges tackled in the work reported in this article.

The semantic pipeline defines the process of transformation and integration of the data from the source format into the final representation i.e. the knowledge graph, see Figure 3. There are two types of knowledge graph creation strategies:

- Materialized Knowledge Graph Creation Process (i.e., data warehousing): In a materialized knowledge graph creation process, data from individual data sources are loaded and materialized into an RDF format and stored in a physical database, the so-called RDF triplestore, e.g., Virtuoso.
- Virtual Knowledge Graph Creation Process (i.e., Data Lake): In a virtual knowledge graph creation process, data remains in the sources (in raw format) and is accessed as needed during query time.

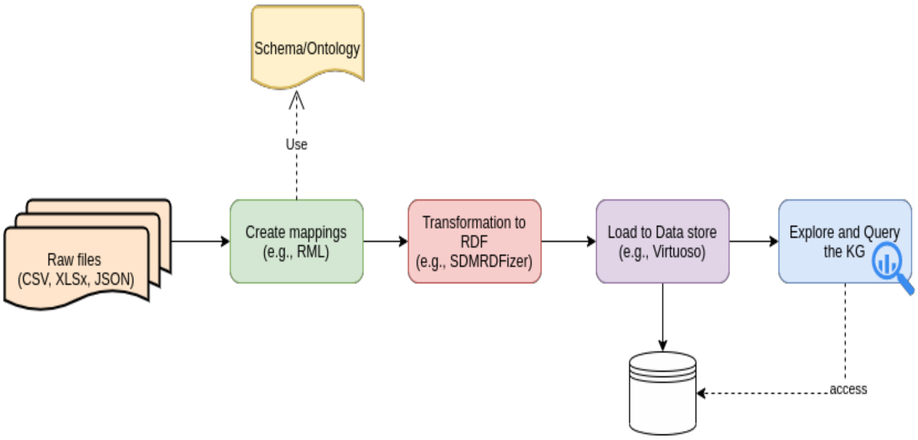

Fig. 3. Semantic pipeline.

We follow the first approach in order to experiment with (1) mechanisms for efficient search and visualization of energy data at different levels of granularity; and (2) provide mechanisms for explainability and interpretability of results of the analytical services. Two mapping languages are utilized for mapping rule definition:

- The Relational to RDF Mapping Language (R2RML) [15] is the W3C recommendation to express customized mapping rules from data in relational databases to generate knowledge graphs represented using the Resource Description Frame work (RDF).
- The RDF Mapping Language (RML) [16] expresses customized mapping rules from heterogeneous data structures, formats and serializations to RDF. RML is a superset of R2RML, keeps the mapping rules as in R2RML but excludes its database-specific references from the core model.

The correspondences among energy data sources and semantic models are described in R2RML and RML. As a result of the execution of R2RML and RML mapping rules, a knowledge graph expressed in RDF is created. Mapping rules are expressed as triples maps. Each triples map refers to a single logical source which can be SQL table or view or data gathered by executing SQL query against the input database. In our case, the mapping rules are applied to transform static data about plants, generation units and weather stations, see Appendix. This data includes geographical location, control area membership and similar data that are not changed frequently. Following examples of the mapping rules focus on PV plants. Since some of the data already exists in a MySQL database, this data is converted to RDF format using the RML-complaint engine, SDM-RDFizer[4]; it executes R2RML and RML mapping rules and transforms raw data in various formats: CSV, JSON, RDB and XML, into an RDF graph knowledge graph. SDM-RDFizer resorts to data structures and physical operators to scale up to large datasets, physical operators, and efficiently execute pipelines of knowledge graph creation [17].

Table 1. The SCADA KG statistics.

Statistics	Value
Total number of RDF triples	18,278,850
Number of classes	83
Number of distinct properties	156
Number of class/subclass pairs	12
Number of different timestamps for timestamped data	1,108,298

Apart from static information about power plants and the grid, measured values from power plants are also collected. The data collected through the SCADA system

---

[4] https://github.com/SDM-TIB/SDM-RDFizer

is available in real-time through a MySQL database; it includes power production forecast, power production measurements, and weather information (e.g., air temperature, wind direction, and solar panel temperature).

The SCADA knowledge graph (KG) is created as a result of execution the mapping rules on top of the MySQL. By the time of this submission, the SCADA KG comprises more than 18M RDF triples, with instances of 83 classes. These classes are described in terms of 156 properties and more than 1M timestamps. Table 1 reports on the characteristics of the current version of the SCADA KG.

4. Knowledge Graph Exploitation

This section presents the services implemented on top of the SCADA KG; they allow for the exploration of the integrated data and their descriptions with the energy semantic data models. SPARQL, the W3C recommendation query language is utilized to express basic queries against the SCADA KG.

4.1 Energy Analytics Dashboard

Since SCADA KG shall work in synergy with various AI-based analytic services and help users to understand results, a visualization tool (EAD - Energy Analytics Dashboard) has been developed. The tool allows fetching data from arbitrary SPARQL end points and supports different analysis / visualization options.

EAD is a data visualization tool that works on top of the SCADA KG. It allows the users to select the data of interest, compare time series (i.e., forecasted load and actual load at that time) and visualize summary statistics on the geographical map. It has been implemented as a web application using JavaScript programming language with help of JQuery library. It uses Highcharts[5] library for visualization, Leaflet[6] library for interacting with geo data. Figure 4 depicts the dashboard and its connection with the pipeline of knowledge graph creation described in Section 3.

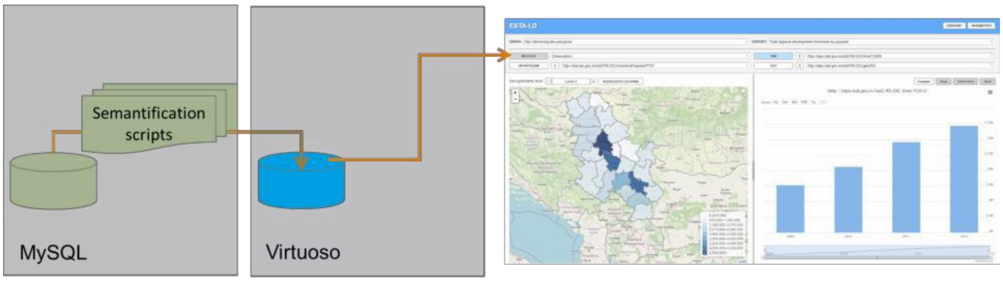

Fig. 4. Semantic pipeline and KG exploration.

---

[5] https://www.highcharts.com

[6] https://leafletjs.com/

4.2 Alignment with EU Initiatives

In November 2012, the CEN-CENELEC-ETSI Smart Grid Coordination Group defined the Smart Grid Reference Architecture [10]. In order to inherently address interoperability, the framework defined five interoperability layers (business, function, information, communication and component layers). The information layer specifies the business context and the semantic understanding. Hence, in future energy smart grids, the technologies described herein are not optional, but mandatory. Currently under development are different energy services marketplaces that in their core include components such as vocabulary management tools and datasets/services registries. In case the production of all PV plants in Serbia can be reached via a SPARQL query, with one click, we can answer the following question "*Show the total energy produced by PV plants in Serbia"*

```
SELECT DISTINCT ?solararray SUM(?value) as ?totalPower
WHERE {
        ?solararray a seas:SolarArray .
        ?solararray art:country <https://projekat-artemis.rs/Country/RS>.
        ?panel seas:isMemberOf ?solararray .
        ?panel a seas:SolarPanel .
        ?panel seas:producedElectricPower ?activePowerProperty .
             }
```

5. Conclusions and Future Work

One of the requirements related to data access procedures in Smart Grids and future electricity markets is the interoperability of energy services. Therefore, this paper proposes an approach for building a knowledge graph enabling semantic interoperability. The semantic data models from the energy sector and the internal SCADA information model are currently used as an information hub materialized in a knowledge graph. It provides the basis for developing and integrating services in the Energy Data Spaces. Additionally, this layer provides the basis for the explainability of machine learning services / analytical applications installed in the smart ecosystem. The future work includes activities that will connect the PUPIN platform with the PLATOON marketplace, thus creating opportunities for broader exploitation of the PUPIN analytical services.

Acknowledgments. The research presented in this paper is partly financed by the EU H2020 funded projects PLATOON (GA No. 872592), the EU project SINERGY (GA No. 952140), and partly by the Ministry of Science and Technological Development of the Republic of Serbia (No. 451-03-9/2021-14/200034) and the Science Fund of the Republic of Serbia (Artemis, No.6527051).

Appendix

```
@base <https://projekat-artemis.rs/> .
<#ARTEMIS_DB> a d2rq:Database;
<#PUPIN_PVPlantMapping> a rr:TriplesMap; rml:logicalSource [
          rml:source <#ARTEMIS_DB>; rr:sqlVersion rr:SQL2008; rml:query """

    SELECT DISTINCT
          plants.id AS plant_id, plants.name AS plant_name, weat-
          er_locations.lat AS lat, weather_locations.lon AS lon, weath-
          er_locations.city AS city, assets.asset_name AS asset_name, coun-
          try.country_code AS ccode,eic_functions.eic_type_function_acronym AS
          eic_func_acronym, organization.organization_short_name AS organiza-
          tion_short_name, organization.organization_name AS organiza-
          tion_name,assets.id AS asset_id
    FROM
          `plants`
    JOIN weather_locations ON plants.weather_location_id = weather_locations.id
    JOIN assets ON plants.asset_id = assets.id
    JOIN organization ON assets.organization_id = organization.id
    WHERE
"""
];
```

References

1. K. Panetta, K.: Five Trends Emerge in the Gartner Hype Cycle for Emerging Technologies 2018," Gartner, 2019. [Online]. https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018, last accessed 2022/03/21.
2. Kiryakov, A.: What is a Knowledge Graph?, 2021. [Online]. Available: https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/, last accessed 2022/03/21.
3. Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., Taylor, J.: Industry-Scale Knowledge Graphs: Lessons and Challenges, Communications of the ACM, vol. 62, no. 8, pp. 36-43, 2019.
4. Jain, S.: Exploiting Knowledge Graphs for Facilitating Product/Service Discovery, arXiv preprint arXiv:2010.05213 (2020).
5. Jain, S., Meyer, V.: Evaluation and refinement of emergency situation ontology. International Journal of Information and Education Technology, 8(10), 713-719 (2018).
6. Janev, V., Vidal, ME., Endris, K., Pujić, D.: Managing Knowledge in Energy Data Spaces. In Companion Proceedings of the Web Conference 2021 (WWW '21 Companion), April 19–23, 2021, Ljubljana, Slovenia. ACM, New York, NY, USA, 11 pages, https://doi.org/10.1145/3442442.3453541, 2021.
7. European Commission: Interoperability solutions for public administrations, businesses, and citizens, https://ec.europa.eu/isa2/, last accessed 2022/03/21.
8. European Commission: The rolling plan on ICT standardisation (2020). https://ec.europa.eu/digital-single-market/en/news/rolling-plan-ict-standardisation, last accessed 2022/03/21.
9. European Commission: European Data Strategy, COM (2020) 66 final. https://ec.europa.eu/digital-single-market/en/policies/75981/3489, last accessed 2022/03/21.
10. CEN-CENELEC-ETS: Smart Grid Reference Architecture, November 2012.
11. Pujic, D., Jelic, M., Tomasevic, N., Batic, M.: Chapter 10 Case Study from the Energy Domain. In Janev, V., Graux, D., Jabeen, H., Sallinger, E. (Eds.) Knowledge Graphs and Big Data Processing. Lecture Notes in Computer Science vol. 12072, pp. 1-208. Springer

International Publishing. ISBN 978-3-030-53198-0. DOI: https://doi.org/10.1007/978-3-030-53199-7_10

12. Berners-Lee, T.: Design issues: Linked Data, http://www.w3.org/DesignIssues/LinkedData.html, last accessed 2022/03/21.
13. Cuenca, J., Larrinaga, F., Curry, E.: DABGEO:A reusable and usable global energy ontology for the energy domain, Web Semantics: Science, Services and Agents on the World Wide Web 61–62 (2020) 100550
14. Lefrançois, M., Kalaoja, J., Ghariani, T., Zimmermann, A.: The SEAS Knowledge Model. ITEA2 12004 Smart Energy Aware Systems Deliverable 2.2, Jan 2017 (PDF - BibTex)
15. Das, S., Sundara, S., Cyganiak, R.: R2RML: RDB to RDF mapping language. In Working group recommendation, World Wide Web Consortium (W3C), 2012.
16. Dimou, A., Vander-Sande, M., Colpaert, P., Verborgh, R., Mannens, E., Van de Walle, R.: RML: a generic language for integrated RDF mappings of heterogeneous data. In Proceedings of the 7th Workshop on Linked Data on the Web, CEUR Workshop Proceedings, 2014.
17. Iglesias, E., Jozashoori, S., Chaves-Fraga, C. D., M. Vidal, M.: SDM-RDFizer: An RML Interpreter for the Efficient Creation of RDF Knowledge Graphs. In The 29th ACM International Conference on Information and Knowledge, 2020.