PAPER • OPEN ACCESS

A semantic data framework to support data-driven demand forecasting

To cite this article: James Allan *et al* 2023 *J. Phys.: Conf. Ser.* **2600** 022001

A semantic data framework to support data-driven demand forecasting

James Allan[1], Francesca Mangili[2], Marco Derboni[2], Luis Gisler[3], Ali Hainoun[4], Andrea Rizzoli[2], Luca Ventriglia[2], Matthias Sulzer[1]

[1] Urban Energy Systems Laboratory, Empa, Ueberlandstrasse 129, 8600 Dübendorf, Switzerland
[2] Dalle Molle Institute for Artificial Intelligence, IDSIA USI/SUPSI, Via la Santa 1, 6962, Lugano, Switzerland
[3] Cividi, Flüelastrasse 10, 8048 Zürich, Switzerland
[4] Digital Resilient Cities, Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria

Abstract. This paper presents a prototype semantic data framework for integrating heterogeneous data inputs for data-driven demand forecasting. This framework will be a core feature of a data exchange platform to improve the access and exchange of data between stakeholders involved in the operation and planning of energy systems. Surveys revealed that these stakeholders require reliable data on expected energy production and consumption for strategic and real-time decision-making. A core feature of the framework is the application of semantic technologies for comprehending spatial and temporal data requirements of energy demand forecasting. This paper demonstrates an approach to meeting these semantic requirements through established data standards and models. The conceptual design process followed the following stages: surveying stakeholders, researching digital technologies' capability, and systematically evaluating the available data. In this paper, we present a prototype based on simulated data. Inputs and results from the simulation model, extracted from open datasets, were structured and stored in a knowledge graph comprised of virtual entities of buildings and geospatial regions. Multiple virtual entities can be linked to a single real-world entity to provide a flexible and adaptable approach to data-driven demand forecasting.

1. Introduction

In Switzerland, the federal council implemented a strategy for a "Digital Switzerland", urging stakeholders from different sectors to implement digital transformation projects [1]. The strategy promotes the application of digital technologies as an opportunity to make the energy industry smarter, flexible and more efficient. The strategy outlines the need for digital tools to link sectors such as mobility and construction to achieve an efficient energy network supplied by sustainable and renewable resources. Digitalization activities are creating a rapid increase in the volume of available data. Opendata.swiss[1] is an initiative and website operated by the Federal Statistical Office where data from different government sectors is published openly and can be used without restriction. The Clean Energy Transition is the shift from fossil fuels to renewable sources. The Swiss Energy Strategy 2050 aims to achieve net zero greenhouse gas emissions by 2050 [2]. Digitalization is viewed as an enabler of achieving climate and environmental targets. A study into approaches for data exchange in the energy sector commissioned by the Swiss Federal Office of Energy provides recommendations for the industry [3]. The report details the highly decentralized nature of data exchange across the numerous stakeholders of the electrical and gas grids. Digital technologies are expected to help grid operators handle the growing complexity and requirements; however, there is a need to improve access and use of data significantly. The report investigates options and critical features to include in a data hub to facilitate

[1] https://opendata.swiss/en/ [Accessed: 01/05/2023]

data exchange in the energy sector. The envisioned data hub will enable: the exchange of metering data, representation of flexibility, external access, implementation of change processes, and end-user offer management. The report also recommends the standardization of data and the use of application interfaces (APIs) to access the data. The report acknowledges that much of the data on the electricity grid operation is standardized according to the Standardised Data Exchange recommendation (SDAT) for the Swiss electricity market approved in 2022.

2. Scope of contribution

This work describes a prototype implementation of a semantic framework that will be integrated into a data-exchange platform to support data-driven energy demand forecasting. The formulation of use cases involved interviewing the various stakeholders of the proposed platform. This paper provides background on concepts critical to the design of the data exchange platform before detailing a prototype framework to handle the semantic requirements of the data.

3. Background

3.1 Data availability and management

Opendata.swiss publishes each dataset with its metadata, including language, spatial & temporal coverage, usage information, and restrictions. The platform, however, allows a vast range of datasets to be published, meaning that they are often heterogeneous in structure, attributes, and formats. For this reason, additional transformation and mapping are required to make the data suitable for applications and link datasets. The datasets published on opendata.swiss are diverse and updated periodically. The structure and contents of published datasets may change between updates, leading to interoperability problems for applications reliant on data. Real-time integration of data from sensors and meters is becoming increasingly important. The rapidly growing and evolving Internet of Things (IoT) sectors enable the interconnection of physical devices, such as sensors and systems, to the Internet, enabling them to collect, exchange and analyze data, opening new capabilities for applications spanning numerous domains [4]. These developments are also relevant for the energy sector in Switzerland, where there is a legally binding target of replacing 80% of conventional meters with smart ones by 2027 [5]. This will provide a source of real-time data for the planning and operation of energy systems. IoT devices such as smart meters can support the implementation of smart grids across all stages of the energy supply chain, from generation to consumption [6]. Despite the growing abundance of data, semantic interoperability between applications such as building energy control, grid operation, and energy system modeling is fundamental to digitalizing the energy infrastructure [7].

3.2 Energy demand forecasting

Utilities and urban planners use demand forecasting for short-term and long-term decision-making in the energy sector. These decisions include:

- **Optimal resource allocation and investment planning:** The energy industry's decision-making process relies on forecasts from a range of time horizons spanning: seconds to hours for demand response, days to months for energy trading, and years or decades for system planning and strategic policy-making [8]. The use of demand forecasting can also be used as inputs to energy system optimization models to support decision-makers [9]. IoT devices provide a data source for demand forecasting [10].
- **Integration of renewable energy resources:** Integrating distributed energy resources can lead to a more holistic modeling of smart grids by understanding how the predicted loads can be met with intermittent generation from renewables [11].

Data-driven methodologies for load forecasting can be classified as simulation- or machine learning-based. Both approaches rely on input data, and the data quality can significantly impact the reliability of the result.

3.2.1 Building energy demand forecasting with physics-based simulation. Accurate building energy simulation is required in the design phase, where there is a need for models that achieve an accurate representation of real-world building operations [12]. Physical models of buildings simulate the energy flows between building components in response to variables such as occupancy and weather. Urban Building Energy Modelling (UBEM) uses less detailed physical models to capture the dynamic and complex interconnections and interdependencies between buildings and their urban environment, e.g. cities and districts [13]. UBEM has fewer data requirements than design stage simulations, and the parameters are often assigned using archetypal approaches [14].

3.2.2 Demand forecasting with machine learning. There is a large number of different objectives and machine learning techniques used for energy planning. These tools are used to directly predict energy demand, production, or the factors that influence it e.g. weather [15]. Load profiles have been used to disaggregate PV production, contributing to a power profile comprising a mix of controllable and uncontrollable loads [16]. A large number of features have been used to train machine learning algorithms to predict the energy demand consumption of buildings (Cooling, heating, and lighting) [17]

3.3 Stakeholder surveys
The platform is designed to improve the access and exchange of data for data providers, need-owners, and digital services. The findings of our surveys revealed that utilities and municipalities rely on data to predict a district's energy consumption and production; digital services, on the other hand, require high-quality data as inputs to their services e.g. system design and optimization [18].

3.4 Review of data architectures and available data
In this paper, data architecture is the collective term given to the processing and structuring of data. This includes using schemas and models to describe the shape and relationships in the data and data formats that specify how the data is encoded. Data architectures were reviewed for specific domains relevant to energy demand forecasting. A summary of data architectures to represent each domain is provided below:

- **Building data.** Information about the construction, materials, geometry, building use, energy supply systems, and fuel type. Building Information Modelling (BIM) for design stage simulation can provide detailed information on buildings. In Switzerland, open datasets such as Swiss Buildings 3D and the Federal Register of Buildings and Dwellings (GWR) provide general information for Urban Building Energy Modelling (UBEM).
- **Spatial boundaries**. Contains georeferenced boundaries for aggregating energy results and linking them to a specific location. The Federal Office of Topology in Switzerland (SwissTopo) publishes the boundaries of administrative regions. H3 is a hierarchical geospatial indexing system for geographic data. Geospatial indexes are not specific to a region and can also be used to aggregate data. Indexed data can be joined across different datasets and aggregated at varying levels of precision. The generic representation of geospatial classes, such as polygons and points, can be achieved using the GeoSPARQL web ontology language.
- **Energy data.** Contains information on energy systems. Data standards such as the Common Information Model (CIM) and FIWARE are designed for the interoperability of smart grids and contain network scale energy concepts. The Brick ontology is a web ontology language used to describe building energy systems.

3.5 Semantic web technologies for energy demand forecasting
Energy demands are the result of a complex combination of societal, economic, technological, and environmental factors. Modellers and engineers have the challenge of incorporating knowledge derived from these domains to improve the accuracy of energy demand forecasts. Semantic web technologies are standards focused on data interoperability, contextual understanding and exchange between knowledge domains. Semantic web technologies have been used to create a common context from large heterogeneous datasets for a building energy stock model [19]. Semantic web technologies have also

 2600 (2023) 022001 doi:10.1088/1742-6596/2600/2/022001

been applied to enrich and reconcile energy performance data [20]. The combination of these features shows that semantic web technologies can improve the accessibility and quality of data for energy demand forecasting.

4. Methodology

4.1 Data preparation and simulation
Energy demands for buildings in the City of Lugano were simulated using the open-source CESAR-P UBEM simulation tool [21]. The inputs for each building were obtained from the Federal Register of Buildings and Dwellings dataset (GWR). The building data was linked to OSM footprint polygons and H3 geospatial cells using "nearest" and "intersection" geospatial joins in QGIS, respectively.

4.2 Semantic mapping
Instance data, including buildings, simulation models, and geospatial cells, were serialized to Turtle format using the RDFLib Python Library and uploaded to a knowledge graph[2]. Instance data was linked to concepts in the Urban Energy Simulation[3] and Digicities[4] ontologies. Both ontologies have been extended for this prototype. The knowledge graph's input and serialized data are published on Zenodo[5]. The ontologies and the relationship between the Virtual Entity and its data sources are shown in Figure 1.

4.3 Machine learning
The knowledge graph was queried to generate structured data (CSV). This data was used as inputs to machine learning models created in CatBoost, LightGBM, and Scikit Learn Python packages to predict aggregated demands using data features attributed to the geospatial indexes.

5. Results & Discussion
This prototype demonstrates how the developed ontologies can organize data for data-driven energy demand forecasting. The aim was to provide a structured and flexible approach to meet data needs and overcome the barriers to data exchange in the energy sector outlined in [3]. The volume of available data for modeling is increasing due to digitalization strategies. This includes real-time and metered data to validate and retrain models to improve forecasting accuracy; however, not all data is open, and mechanisms to establish data agreements, ensure privacy and reduce bureaucratic barriers are still required [22]. Nevertheless, a data framework that evolves and adapts to meet the specific needs of demand forecasting is required. We introduce the Virtual Entity as a class of the Digicities ontology that can be linked to a Physical Record of a real-world entity using the PROV-O ontology. The Virtual Entity can be any real-world entity, such as a building or a geospatial region. This representation allows multiple Virtual Entities to be linked to a single Physical Record. Linking demands to a Virtual Entity enables approaches to be compared and improved as data becomes available. In the preliminary study, machine learning models were trained on data instances to predict the simulated demands for individual buildings and geospatial cell aggregations. At the aggregated level, additional features were engineered, including total surface area, building count, and the minimum, maximum, and average number of floors. The machine learning model achieved similar coefficients of determination (R2) between 0.86 and 0.94 for single buildings and 0.87 and 0.97 for data geospatial aggregations based on the H3 indexing. The similarity in these results is due to the models being trained on features directly linked to the simulation inputs. Therefore, these results are demonstrative and future work will involve integrating data from various sources. This framework will be expanded as part of a data exchange platform to integrate actual data, simulated data, and machine learning models to provide insights for decision-makers. Future machine learning models will predict the performance of metered data using data features linked to the description of virtual entities stored on the platform. In addition to engineering new features from the

---
[2] https://graphdb.nestcloud.ch/ [Accessed: 01/05/2023]
[3] https://ja98.github.io/ues [Accessed: 01/05/2023]
[4] https://ja98.github.io/digicities_ontology/ [Accessed: 01/05/2023]
[5] https://doi.org/10.5281/zenodo.7875922 [Published: 12/05/2023]

data, machine learning models will improve data quality for entities using classification. Integrating these features aims to reduce the time to identify and process data for machine learning models. In addition, semantic web technologies will increase the quantity and quality of data used for machine learning and simulation. This will enable the models to consider additional training features and improve the accuracy of the forecasts. These metrics will be evaluated across the pilot regions of this study. The underlying semantic framework will also be extended to facilitate the data exchange process by connecting to specific clauses of data usage agreements. This includes specific restrictions on data linking and reporting outputs to protect privacy.

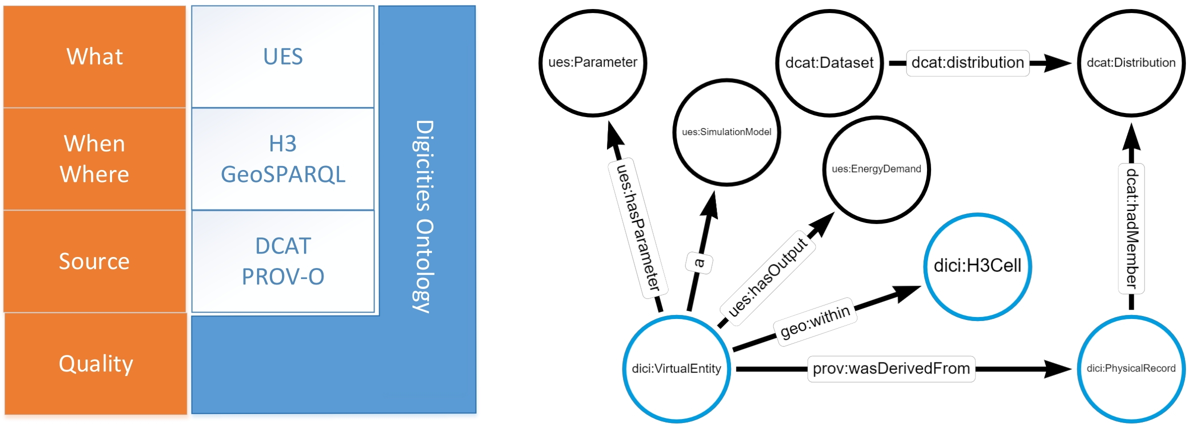

Figure 1. Left – The semantic requirements and the data architectures used fulfil them. Right – An overview of the semantic data structures that trace assumptions to the original data sources. The blue nodes show entities introduced in this work. The prefixes indicate the use of ontology namespaces.

6. Conclusion

This paper presents a semantic framework to support data-driven energy demand forecasting. We have outlined how semantic requirements can be achieved using a combination of established ontologies. We have shown how the semantic framework can support simulation and machine-learning approaches to demand forecasting. In the next phase, the framework will be integrated and expanded as part of a data exchange platform, where the objective is to improve interoperability, data discovery, and data quality. We will work closely with stakeholders, including data providers, need-owners, and digital services, to integrate additional data types and validate across use cases.

Acknowledgements

This project received support from the Swiss Federal Office of Energy through the framework of the joint programming initiative ERA-Net Smart Energy Systems' focus initiative Digital Transformation for the Energy Transition, with support from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883973.

Credit-author statement

James Allan: Conceptualisation, Writing - Original Draft, Project administration, Data Curation, Visualization, Supervision, Funding acquisition. **Francesca Mangili**: Conceptualisation, Writing - Original Draft, Formal analysis. **Marco Derboni:** Conceptualisation, Writing - Review & Editing. **Luis Gisler:** Conceptualisation, Funding acquisition. **Ali Hainoun:** Conceptualisation, Funding acquisition. **Andrea Rizzoli:** Conceptualisation, Funding acquisition, Writing - Review & Editing. **Luca Ventriglia**: Formal analysis. **Matthias Sulzer:** Conceptualisation, Funding acquisition.

References

[1] BFS 2020 Digital Switzerland Strategy
[2] Swiss Federal Office of Energy 2022 Energy Strategy 2050

[3] Holles S, Rütten F, Türkucar T and Oliver Thier C 2021 *Datahub Schweiz.* (Bern: Bundesamt für Energie)
[4] Gubbi J, Buyya R, Marusic S and Palaniswami M 2013 Internet of Things (IoT): A vision, architectural elements, and future directions *Future Gener. Comput. Syst.* **29** 1645–60
[5] Schweizerische Bundesrat 2008 *Stromversorgungsverordnung*
[6] Ghasempour A 2019 Internet of things in smart grid: Architecture, applications, services, key technologies, and challenges *Inventions* **4** 22
[7] Pritoni, Marco, et al. "Metadata schemas and ontologies for building energy applications: A critical review and use case analysis." Energies 14.7 (2021): 2024.
[8] Hong T and Fan S 2016 Probabilistic electric load forecasting: A tutorial review *Int. J. Forecast.* **32** 914–38
[9] Scheller F and Bruckner T 2019 Energy system optimization at the municipal level: An analysis of modeling approaches and challenges *Renew. Sustain. Energy Rev.* **105** 444–61
[10] Li L, Ota K and Dong M 2017 When weather matters: IoT-based electrical load forecasting for smart grid *IEEE Commun. Mag.* **55** 46–51
[11] Habbak H, Mahmoud M, Metwally K, Fouda M M and Ibrahem M I 2023 Load Forecasting Techniques and Their Applications in Smart Grids *Energies* **16** 1480
[12] Coakley D, Raftery P and Keane M 2014 A review of methods to match building energy simulation models to measured data *Renew. Sustain. Energy Rev.* **37** 123–41
[13] Hong T, Chen Y, Luo X, Luo N and Lee S H 2020 Ten questions on urban building energy modeling *Build. Environ.* **168** 106508
[14] Wang D, Landolt J, Mavromatidis G, Orehounig K and Carmeliet J 2018 CESAR: A bottom-up building stock modelling tool for Switzerland to address sustainable energy transformation strategies *Energy Build.* **169** 9–26
[15] Debnath K B and Mourshed M 2018 Forecasting methods in energy planning models *Renew. Sustain. Energy Rev.* **88** 297–325
[16] Salani M, Derboni M, Rivola D, Medici V, Nespoli L, Rosato F and Rizzoli A E 2020 Non intrusive load monitoring for demand side management *Energy Inform.* **3** 25
[17] Amasyali K and El-Gohary N M 2018 A review of data-driven building energy consumption prediction studies *Renew. Sustain. Energy Rev.* **81** 1192–205
[18] Bollinger L A and Dorer V 2017 The Ehub Modeling Tool: A flexible software package for district energy system optimization *Energy Procedia* **122** 541–6
[19] Hoare C, AlQazzaz T B M, Ali U, Hu S and O'Donnell J 2023 Development of a National Scale Digital Twin for Domestic Building Stock *LDAC2023*
[20] Popa A, Ramallo González A P, Jaglan G and Fensel A 2022 A Semantically Data-Driven Classification Framework for Energy Consumption in Buildings *Energies* **15** 3155
[21] Orehounig K, Fierz L, Allan J, Eggimann S, Vulic N and Bojarski A 2022 CESAR-P: A dynamic urban building energy simulationtool *J. Open Source Softw.* **7** 4261
[22] von Grafenstein, Max 2022 Reconciling Conflicting Interests in Data through Data Governance. An Analytical Framework (and a Brief Discussion of the Data Governance Act Draft, the Data Act Draft, the AI Regulation Draft, as well as the GDPR)