Big Data Pipeline for Building Energy Management

Zhiyu Pan [1,*], Panagiotis Kapsalis [2], Konstantinos Alexakis [2], Georgios Korbakis [2], and Antonello Monti [3]

[1] Institute for Automation of Complex Power Systems, RWTH Aachen University, Aachen, Germany;
[2] Decision Support Systems Laboratory, National Technical University of Athens, Athens, Greece;
Email: pkapsalis@epu.ntua.gr (P.K.), kalexakis@epu.ntua.gr (K.A.), gkorbakis@epu.ntua.gr (G.K.)
[3] Institute for Automation of Complex Power Systems, RWTH Aachen University, Aachen, Germany;
Email: amonti@eonerc.rwth-aachen.de (A.M.)
* Correspondence: zhiyu.pan@eonerc.rwth-aachen.de (Z.P.)

***Abstract*—The increasing of heterogeneous data in the building domain brings a huge challenge to data integration. With the combination of ontology and data model, a building energy domain common data model is developed and provides a uniform data schema to guide the data integration process. Additionally, a cloud data pipeline is proposed and developed, which includes the common data model, data harmonization, data storage and data querying. The requirement and possible use cases for the big data pipeline for building energy management are described. This work provides guidelines for big data management in building energy domain. Furthermore, our data pipeline is evaluated with 11 large pilots and shows a significant improvement in the data governance process.**

***Keywords*—big data, building Life-Cycle, data model, data pipeline, ontology**

I. Introduction

The explosive increase of IoT devices produces huge heterogeneous data in the building domain. According to (Jiang *et al.*, 2016), the increase of the volume in big data also raises the problem in data storage, querying and processing. Traditional approaches to dealing with data heterogeneity have depended on the use of data models. The problem of data model has been pointed out in reaching an agreement among a community of users, as well as the data models' lack of flexibility in adapting to changes and the loss of information after exporting and importing data through apps. The authors of (Noura *et al.*, 2019) point out that the harmonization effort of the huge heterogeneous data is significantly decreased with ontology. Ontology is an explicit specification of a conceptualization, which provides a shared vocabulary and the relationships among them across the internet. The important feature is the information system from closed and stand-alone to distributed, loosely coupled systems through ontology. A plethora of ontologies for many applications have been produced. However, under the condition of big data, ontologies are huge in terms of entity and relationships that make the search for the most suitable ontology difficult and decrease the reusability (Caldarola and Rinaldi, 2016). Furthermore, ontologies for building energy management are too general and fragmented to be useful in practice, and the current ontologies are not flexible enough to include innovative sensors, such as a Kindle or an Amazon Echo (Bhattacharya *et al.*, 2015). The goal of data model is to structure the task-oriented information, while the ontology provides the generic representation of data (Spyns *et al.*, 2002). Compensating ontology with data model, the data model is expanded and provides the developer explicit domain knowledge. Therefore, this work proposes a common data model, which combines the ontology and data model and provides the understanding of the members of the community and helps to decrease ambiguity in communication. Another issue in the building energy domain is that there is no single ontology, which covers the whole building life-cycle (Ramesh *et al.*, 2010): Manufacturing Phase, Use Phase and Demolition Phase; and considers from different perspectives and scales of building: buildings as individual elements (building level) to their aggregation at various scales (district to national level).

The main issue with the storage of building energy information is that the static and real-time data are kept in different places and different formats (e.g., semi-structured and structured data). The larger the volume of data is increased, the larger the possibility of error having data misstated with different types and formats (Marinakis *et al.*, 2020). Moreover, no data lake or data warehouse collects all data related to energy information of buildings. This limits the data access and availability for energy analysts. With relational and non-relational databases that host the data applicable for analysis, it is a challenging task to query all the information and data with different formats and types. Therefore, the need for data querying has emerged.

The contributions of this paper are: 1. To solve all the problems motioned before, this work proposes a big data

Manuscript received June 2, 2023; revised August 22, 2023; accepted October 5, 2023.
Copyright credit, project number: 1010000158

pipeline for building energy management based on the reference architecture in (Pau *et al.*, 2022) that enables the flexible handling of building energy big data from applications or sources and demonstrate the pipeline with data from 11 pilots and explicit implementation. 2. To solve the problem of heterogeneous data, this paper introduces a new common data model: Building Energy Domain Common Data Model, which reuses not only the existing ontology (e.g. Brick (Balaji *et al.*, 2018) and SAREF (DanieleDaniele *et al.*, 2015)) but also the data model (e.g. FIWARE (https://www.fiware.org/ smart-data-models/) and EPC4EU (Serna-Gonz´alez, *et al.*, 2021)). Additionally, our data model covers the whole building life cycle and different perspectives and scales of building. 3. The storage of the harmonized information to a database technology that supports embedded formats and the querying of data relying on different sources is implemented. To control the amount of data and manage in-memory processes the proposed solution is built upon a data warehouse system that receives streams of data from applications and sources, store them in a document-oriented database and enables the querying of stored data from the upper layers by using a simple query language like SQL. Moreover, an enriched warehouse of harmonized building data is structured that receives real-time and batch data via its streaming & batch mechanisms. The harmonization process is enabled via the construction of a Common Data Model that polishes the incoming data and structures them in a schema-less database. The data are stored in different collections that are named from the topics that data are queued. A querying engine is on top to query and combines the stored building information.

II. CONCEPT

A. Use Cases and System Requirements

Big data and associated technologies are gaining traction, creating an unprecedented potential to improve energy efficiency across the building sector and life cycle, as well as better manage energy use and generation at the building level. This has aided in the transformation of buildings into digitally upgraded edge hubs capable of successfully managing and controlling their energy generation and consumption while engaging with other smart energy components of the future energy system. With properly energy data analysis, stakeholders benefit with more comprehensive actionable insights, as well as improve decision-making. A variety of data analysis techniques (including among others optimization, forecasting, classification and clustering) can be applied to the aforementioned amounts of big data and on top of our data pipeline, supporting the design of new data-driven business models for several beneficiaries, such as national and local governments, network operators and suppliers, Energy Service Companies (ESCOs), building managers and facilitators, construction and renovation sector, investors and financiers, policy makers, and researchers.

The use cases, which are possible to build on top of our data pipeline, are summarized in four categories.

1. Performance: analytics for energy performance based on the operational stage of buildings aimed at monitoring and improving their energy performance. Predictive capabilities related to comfort evaluation, energy demand, consumption or generation, will be complemented by optimization capabilities for the management of comfort-aware building energy consumption. To conceptualize, all the device data models related to building energy are required. SAREF and SAREF4BLDG were created for the concept of IoT devices especially for building domain, which fulfils the requirement of this use case.

2. Design: the design category is to facilitate the design, refurbishment and development of building infrastructure. In particular, this use case focuses on building level design of retrofitting actions and on district level design of networks.

3. Policy: this use case is to support policymaking and policy impact assessment. They will be targeting three main elements revolving around policies at different scales: Sustainable Energy and Climate Action Plans, Energy Performance Certificates (EPC) and impact assessment of EU policies for buildings. Therefore, EPC and weather data model are included in the Common Data Model. The EPC is variant across the whole European, which is also considered in our Common Data Model.

4. Fund: this fund use case aims to perform finer-grained prediction for building comforts by integrating a variety of historical data on energy efficiency investments with near real-time metered energy consumption, thus contributing to better define Energy Performance Contract conditions. It is tailored to ESCOs and financing institutions. This use case also addresses the centralization of building stock data, and the analysis of refurbishment actions.

Therefore, the data Pipeline should fulfil the following requirement. First, it should provide a well-defined API for data export. Second, the common data model should cover the Real-time data generated by IoT technologies (e.g. energy consumption and energy production using smart meters, sensor-based data); historical data for model development and pattern recognition (e.g. old weather data, energy costs); Open data from various sources (weather, climate, EPCs SECAPs, costs); Secondary data not related directly to energy and climate (e.g. demographics, economics, cadaster). Thirdly, datasets in different formats and sizes are able to be imported into the pipeline. Last, the pipeline should provide optimal space allocation in the data storage.

B. Architecture of Cloud Data Pipeline

The big data lifecycle is defined in four different phases in (Hu *et al.*, 2014): data generation, data acquisition, data storage, and data analytics. The data acquisition phase consists of data collection, data transmission and data pre-processing. This paper focuses on data acquisition and data storage part of big data phases to provide the middleware of big data life-cycle especially for building energy domain.

The pipeline is structured into five parts in Fig. 1: data collection, data prepossessing service, data harmonization, storage and querying. The Kafka (Goodhope *et al.*, 2012) is an open source streaming processing system, which can handle real-time data, commonly at the second or even millisecond level. It is distributed and scalable and offers high throughput. Therefore, Kafka is used to collect datasets from different databases and transport those to the data prepossessing service. This data prepossessing service focuses mainly on the data cleansing technique to determine inaccurate, incomplete, or unreasonable data. In (Truong *et al.*, 2020) and (Rekatsinas *et al.*, 2017), several methodologies used in data cleansing process are listed, which is possible to integrate in this module. Afterward, the data harmonization, which is aim to covert the data into a unified view, is executed according to the common data model. MongoDB database is selected to store the harmonized data, which facilitates the storage of both structured and unstructured data collections and can manage a high volume of data loads. For the implementation of the database-agnostic data warehouse, which means to query different databases, relational and non-relational, by using SQL the PRESTO functionalities. PRESTO is a distributed SQL query engine that provides interactive workloads by querying many different data sources. In this case, PRESTO is configured on top of MongoDB, enabling with this way the big data querying and analysis of the harmonized building data over a memory-based architecture without moving the aforementioned datasets to another structured system. Additionally, the data-querying layer provides APIs for the different services (e.g., energy performance monitoring, support policy making and policy impact assessment).

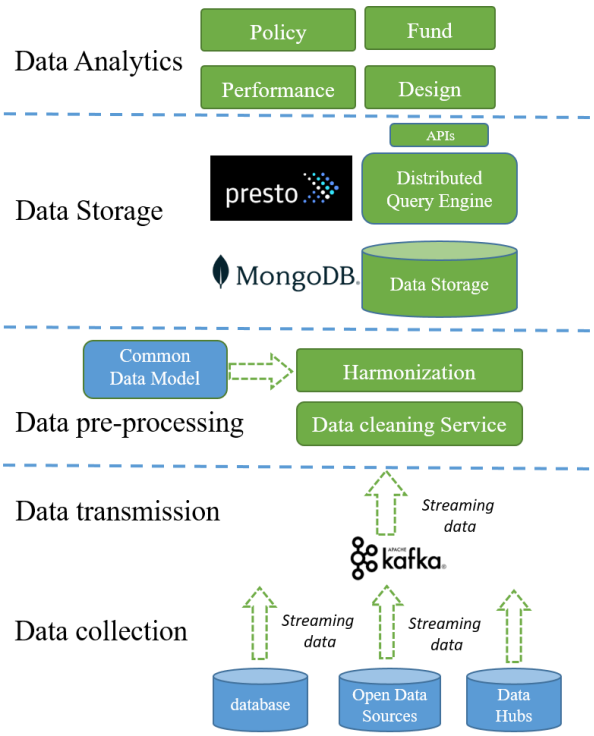

Figure 1. Architecture of cloud pipeline.

C. Building Energy Domain Common Data Model

To create a building domain common data model, the ontology development 101 method (Noy *et al.*, 2001) is applied. It is the most highly cited method compared with all other ontology development methods and provides a clear guideline for the ontology development with the most popular and widely used ontology tool Protégé. In (Constantin *et al.*, 2017), the ontology development 101 method is applied to combine building standards and power systems digital automation standards into one ontology for building energy management and a data model is developed afterward. This provides the inspiration for how to apply the ontology development method to develop Common Data Model and combine the data model and ontology together. This ontology development method is modified to develop data model and enable the functionality of reusing the existing ontology and data model, which is summarized as follows:

- Determine the domain and scope of the data model
- Determine possible reuse of existing ontology and data model
- Enumerate important terms
- Define the classes and the class hierarchy
- Define the properties of classes-slots
- Define the facets of the slots
- Create instances
- Convert Ontology to data model

The first step about the domain and scope has already been described in the section on system requirements. The existing ontology and data models related to the building energy domain are described below. FIWARE smart data model provides different data models in different smart domains. Each entity contains enriched attributes, which provide the user with high coverage of the vocabulary. The smart cities and smart energy domains data model are especially important in the building energy domain. However, FIWARE smart data model lacks the hierarchy between smart cities and smart energy domains and the class hierarchy in the smart city domain. It only provides the definition of the classes and the properties and the facets of the slots. Brick ontology is a uniform metadata schema for representing buildings and contains sensors, subsystems and the relationships between them. The limitation of Brick is that the detail of each class is missing (e.g., the properties and the facets). SAREF is an ontology, which identifies 20 recurring concepts from different ontologies in the home and buildings domain. Except for the core ontology, there are different extensions in some domains, which can provide many relevant domain-specific features and the modelers with less experience and explicit domain knowledge (Pritoni *et al.*, 2021). (Pritoni *et al.*, 2021) points out that the represent geometry and functions of spaces and control strategies for control devices are not included in all the SAREF extensions. EPC4EU data model aims to model the Energy Performance Certificate (EPC) datasets at different geographical scales from local to European level. The use of the EPC4EU data model allows the comparison of EPC datasets across European and

supports energy efficiency policies making. This is exactly the part of the goal in our building energy domain common data model. However, the EPC4EU includes only the building and EPC information, without the sensors and subsystems. The last step is to translate the developed ontology to data model with the method in (Trinkunas and Vasilecas, 2007), which provides a graph-oriented method for ontology transformation into data model. The detail of the whole data model is not described in this paper and the top-level structure of the data model is illustrated in the result section.

D. Data Harmonization

In the literature (Kumar *et al.*, 2021), the state of the art data harmonization techniques is analyzed. The result shows that the heterogeneity of structured, semi-structured and unstructured data is managed by using Natural Language Prepossessing (NLP), machine learning, deep learning and ontology technology. The method from (Hong *et al.*, 2019) (Pan *et al.*, 2022) try to solve the heterogeneity problem applying NLP and ontology technology, which shows a significantly improvement in the efficiency compared with the manually harmonization. Therefore, this work adopts this method in our data harmonization module, which is summarized in the following steps: define the input data, define the mapping rules, convert the attribute according to mapping rules, convert the value of the attribute according to the data model, and generate the output.

E. Data Storage

Building data are captured from sensors and IoT devices installed in buildings. Systems called Building Management Systems (BMS) are responsible for building data capturing and storage. The indicated real-time conditions (e.g., energy consumption, indoor humidity and temperature) are stored in relational databases such as PostgreSQL and MariaDB for later processing (Marinakis and Doukas, 2018). The advantage of this solution is its simplicity as the end users of the storage need to utilize the SQL language to retrieve the stored data. The main drawback is that relational databases are standalone and in case of a database, downtime the data will be lost, another approach that improves the storage problem is presented in this research (Marinakis *et al.*, 2020) where the building data are stored in a flat format suitable for fast querying. The type of database used is a NoSQL Hadoop cluster, which is the main storage solution. MapReduce operations are applied over stored building flat schemas and simultaneously this solution provides high availability, speed, and scalability and fault tolerance. The main issue occurred with the aforementioned storage architecture are usage problems from storage end users due to the complexity of MapReduce and the non-existence of a unified data schema that standardizes the operations over stored data.

The next step of the cloud pipeline, after the data harmonization procedure, is the storage of the LSPs datasets. Due to the nature of semi-structured data the selected database technology should support nested formats with multiple properties, with variations on schemes of the stored datasets. For that reason, the MongoDB NoSQL database is selected. As a document-oriented database, MongoDB facilitates the storage of both structured and unstructured data collections and can manage high volume of data loads. In the cloud pipeline, the harmonized datasets are consumed from Storage Kafka consumer and then persisted in document collections. For each different Large-Scale Pilot (LSPs), a MongoDB collection accommodates its data load.

F. Data Querying

Data querying solutions for building energy data are focused only on metadata querying and insights extraction. The proposed querying architecture in this research (Kapsalis *et al.*, 2022) leverages a graph database that receives batch data in JSON format and then transforms them into graph entities. Furthermore, the graph database persists ontologies and RDF (Resource Description Framework Schema) to enhance the stored metadata. By leveraging stored metadata patterns and relationships a REST API on top of the graph database receives JSON input and returns results from stored metadata. The main drawback of this data querying architecture is that takes into consideration only building metadata. The building energy management systems expose time-series data from sensors that measure (e.g., temperature, humidity, and produced energy) and there is the need to manage this type of data (Marinakis *et al.*, 2013). Nowadays enriched warehouses are multi-database systems that handle metadata databases and time-series databases. The emerging need is the management and querying of both data stores (real-time and metadata databases) and combining the stored information, A system that is database agnostic is needed to hide each database query language and the end user will be capable to write SQL to query the stored building information.

In general, the goal of the cloud pipeline, in terms of storage, is to build an enriched data warehouse where harmonized building energy data are collected and queried via intelligent procedures and periodic tasks with an agnostic database manner. Furthermore, this data warehouse will provide the possibility to fuse external datasets with the stored LSPs data. For the implementation of the database-agnostic data warehouse, which means querying different databases, relational and non-relational, by using SQL the PRESTO query engine functionalities. PRESTO is a distributed SQL query engine that provides interactive workloads by querying many different data sources. In this case, PRESTO is installed and configured on top of MongoDB, enabling with this way the big data querying and analysis of the harmonized building data over a memory-based architecture and without moving the aforementioned datasets to another structured system.

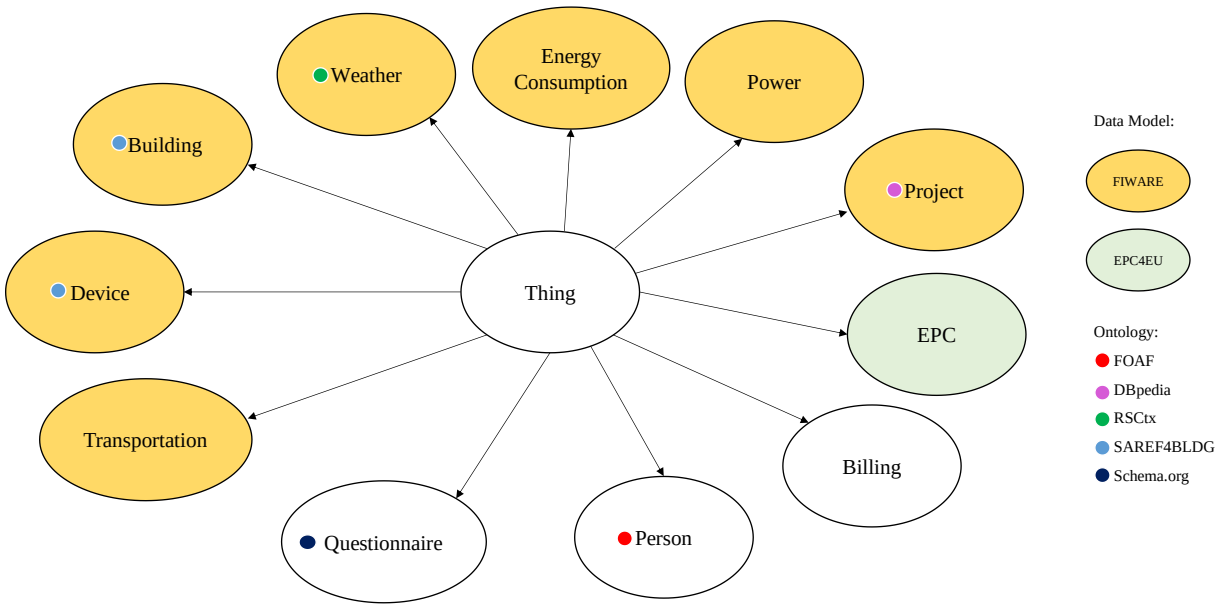

Figure 2. Building domain common data model.

III. DEMONSTRATION

The data pipeline is evaluated with 11 pilots across nine countries, which cover the whole building life cycle and different perspectives and scales of building in MATRYCS project. The workflow of the architecture will be demonstrated how to use the Common Data Model to guide the data harmonization, store the harmonized data and how to query it afterward.

The building domain common data model is developed to cover the completely building life cycle and different perspectives and scales of building in the 11 pilots and is divided into ten categories in Fig. 2: Building, Device, Energy Performance Certificate, Energy Consumption, Transportation, Project, Person, Questionnaire, Billing and Weather. Each category contains different classes and each class contains different attributes. Those categories are based on the Fiware smart data model, which separates the data model according to different smart domains, each smart domain contains different categories and each category contains different classes. Because the focus for this study is on building energy domain. The Common Data Model is not separated at the domain level but at the category level. Based on those level definitions, it provides the modularity and simple extendibility with other categories or even other domains. The detail is not illustrated in Fig. 2. The legend on the right side shows what kind of data model and ontology is used in the common data model.

In the following, the developed data pipeline is demonstrated whit LSP 1 BTC Tower data. The first step is harmonization the BTC Tower data according to the common data model. The sample harmonization process is illustrated in Fig. 3 The upper table refer to the raw data with CSV format as input. The bottom half of the picture shows the output in JSON-LD format. The harmonization module according to NGSI-LD standard (https://www.etsi.org/deliver/etsigs/CIM/001 099/009/01.01. 01 60/gs cim009v01010), which developed by the European Telecommunication Standards Institute (ETSI) and used in FIWARE smart data model, creates the properties type and id automatically.

DATE	TIMESTAMP	LOCATION	ENERGY_ SOURCE	MEASURE	UNIT_OF_ MEASURE	INTERVAL	VALUE
12/1/ 2020	2020-12-01 00:00:00:000	BTC tower	Electricity	billing meter total consumption	kWh	H	106.4

Harmonization


```
{
    "dateObserved": 1606780800000,
    "address": "BTC tower",
    "controlledProperty": "Electricity",
    "source": "billing meter total consumption",
    "supportedUnit": "kWh",
    "value": 3000.0,
    "type": "Device",
    "id": "Device-BTC_tower"
}
```


Figure 3. Sample harmonization process.

The harmonized BTC Tower data are send through a BTC Tower Kafka topic and then stored by leveraging Storage mechanism procedures to a BTC Tower MongoDB collection. Fig. 4 depicts how the data are stored and structured in MongoDB. Data Storage is consisted of two sub-components. The first sub-component is the Kafka Storage Consumer, which receives the events from topics where the consumer is subscribed. When an incoming event from a topic is received from Storage Consumer, it is stored in MongoDB, which is the second component of MATRYCS Data Storage mechanism. Kafka Storage Consumer leverages pymongo, which is the official MongoDB Python driver to persist the new information to MongoDB collections. Each topic has a respective collection in MongoDB, which means that messages originated form topic X are stored in collection X.

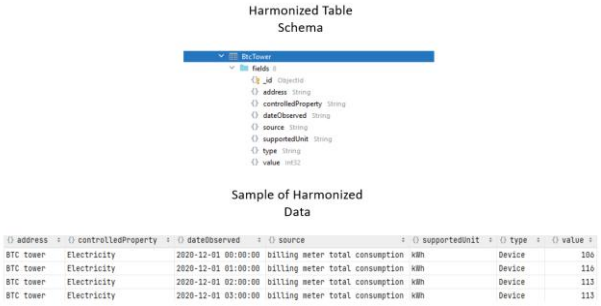

Figure 4. Harmonized BTC tower MongoDB collection.

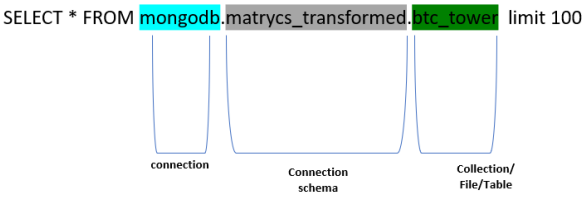

Figure 5. BTC tower querying via query engine.

After their insertion in MATRYCS Storage the data needs to be exposed in MATRYCS applications, that consists the MATRYCS upper layer where the data are used for models training. Our contribution in this part is the addition of a query engine that ensures the query abstraction from upper layers and the possibility to add more databases and external data for combining them and perform joins over these data and other relational operations. The query engine is connected with relational and non-relational structures and the end user needs to write SQL queries to access the stored MATRYCS large-scale pilot data. In MATRYCS case, the query engine is connected with MongoDB and SQL queries are applied from MATRYCS Analytics layer to query the stored data. In order to demonstrate the functionality, this study uses the LSP 1 BTC Tower dataset and the following query that is presented bellow returns the stored BTC Tower data (Fig. 5). It is obligatory to define the connection name, which is the connected structure, the table schema, which is the table schema where tables and connections exist. Finally, it is needed to be defined the table and collection name to query. The query bellow fetches all data stored in BTC Tower limit 100.

IV. Conclusion

Drawing on the building energy reference architecture, this study has demonstrated the big data pipeline for building energy management and focused on the big data variety problem. The pipeline provides a guideline and reference implementation, which also improves the interoperability between different data sources using ontology. Each step of the data pipeline is discussed and analyzed through the LSPs data. Especially the Building Energy Domain Common Data Model, which covers the whole building life cycle and different perspectives and scales of building, is developed based on the existing ontology and data model. In the current version of data pipeline, this work only processes the semi-structured and structured data as input, which will be extended to cover unstructured in the future version.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Zhiyu Pan and Antonello Monti conducted the research; Konstantinos Alexakis and Georgios Korbakis analyzed the data; Zhiyu and Panagiotis Kapsalis wrote the paper; all authors had approved the final version.

Acknowledgment

The authors would like to thank all the MATRYCS consortium partners and especially BTC & Coopernico for the data and discussions during the implementation.

References

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., et al., 2018. Brick: Metadata schema for portable smart building applications. ***Applied energy***, 226: 1273–1292.

Bhattacharya, A., Ploennigs, J. & Culler, D. 2015. Short paper: Analyzing metadata schemas for buildings: The good, the bad, and the ugly. ***Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments***, pp. 33–34.

Caldarola G. E. & Rinaldi, M. A. 2016. An approach to ontology integration for ontology reuse. ***Proceedings of 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)***, pp. 384–393.

Constantin, A., L¨owen, A., Ponci, F., M¨uller, D. & Monti, A. 2017. Design, implementation and demonstration of embedded agents for energy man- agement in non-residential buildings. ***Energies***, 10(8): 1106.

Daniele, L., Hartog, d. F. & Roes, J. 2015. ***Created in close interaction with the industry: the smart appliances reference (saref) ontology.*** Paper presented at International Workshop Formal Ontologies Meet Industries, pp. 100–112.

Goodhope, K., Koshy, J., Kreps, J., Narkhede, N., Park, R., Rao, J. & Ye, Y. V. 2012. Building linkedin's real-time activity data pipeline. IEEE Data Eng. ***Bull***., 35(2): 33–45.

Hong, N., Wen, A., Shen, F., Sohn, S., Wang, C., Liu, H. & Jiang, G. 2019. Developing a scalable fhir-based clinical data normalization pipeline for standardizing and integrating unstructured and structured electronic health record data. ***Jamia Open,*** 2(4): 570–579.

Hu, H., Wen, Y., Chua, -S. T. & Li, X. 2014. Toward scalable systems for big data analytics: A technology tutorial. ***IEEE access***, 2: 652–687.

Jiang, H., Wang, K., Wang, Y., Gao, M. & Zhang, Y. 2016. Energy Big Data: A Survey. ***IEEE Access,*** 4: 3844–3861.

Kapsalis, P., Kormpakis, G., Alexakis, K. & Askounis, D. 2022. Leveraging graph analytics for energy efficiency certificates. ***Energies,*** vol. 15, no. 4. Available: https://www.mdpi.com/1996-1073/15/4/1500

Kumar, G., Basri, S., Imam, A. A., Khowaja, A. S., Capretz, F. L. & Balogun, O. A. 2021. Data harmonization for heterogeneous datasets: A systematic literature review. ***Applied Sciences***, 11(17): 8275.

Marinakis, V., Doukas, H., Tsapelas, J., Mouzakitis,´S., Sicilia, A.,

Madrazo, L. & Sgouridis, S. 2020. From big data to smart energy services: An application for intelligent energy management. ***Future Generation Computer Systems***, 110: 572–586.

Marinakis V. & Doukas, H. 2018. An advanced iot-based system for intelligent energy management in buildings. ***Sensors,*** 18(2). Available: https://www.mdpi.com/1424-8220/18/2/610

Marinakis, V., Doukas, H., Tsapelas, J., Mouzakitis, S., Alvaro Sicilia, Madrazo, L. & Sgouridis, S. 2020. From big data to smart energy services: An application for intelligent energy management. ***Future Generation Computer Systems,*** 110: 572586. Available: https://www.sciencedirect.com/science/article/pii/S0167739X17318769

Marinakis, V., Karakosta, C., Doukas, H., Androulaki, S. & Psarras, J. 2013. A building automation and control tool for remote and real time monitoring of energy consumption. ***Sustainable Cities and Society***, vol. 6, pp. 11–15. Available: https://www.sciencedirect.com/science/article/pii/S2210670712000467

Noura, M., Atiquzzaman, M. & Gaedke, M. 2019. Interoperability in internet of things: Taxonomies and open challenges. ***Mobile networks and applications***, 24(3): 796–809.

Noy, F. N., McGuinness L. D. *et al.*, 2001. Ontology development 101: A guide to creating your first ontology.

Pan, Z., Pan, G., & Monti, A., 2022. Semantic-Similarity-Based Schema Matching for Management of Building Energy Data. ***Energies***, 15(23): 8894.

Pau, M., Kapsalis, P., Pan, Z., Korbakis, G., Pellegrino, D. & Monti, A. 2022. Matrycs—a big data architecture for advanced services in the building domain. ***Energies***, 15(7): 2568.

Pritoni, M., Paine, D., Fierro, G., Mosiman, C., Poplawski, M., Saha, A., Bender, J. & Granderson, J. 2021. Metadata schemas and ontologies for building energy applications: A critical review and use case analysis. ***Energies***, 14(7): 2024.

Ramesh, T., Prakash, R. & Shukla, K. 2010. Life cycle energy analysis of buildings: An overview. ***Energy and buildings***, 42(10): 1592–1600.

Rekatsinas, T., Chu, Ilyas, F. I., & R´e, C. 2017. Holoclean: Holistic data repairs with probabilistic inference. ***arXiv preprint arXiv***, 1702.00820.

Serna-Gonz´alez, V., Hern´andez Moral, G., Miguel-Herrero, F., Valmaseda, C., Martirano, G., Pignatelli, F. & Vinci, F. 2021. Elise energy & location applications: Use case "harmonisation of energy performance certificates of buildings datasets across eu – final report. ***Luxembourg***. Available: https://www.etsi.org/deliver/etsigs/CIM/001099/009/01.01.0160/gscim009v010101p.pdf

Spyns, P., Meersman, R. & Jarrar, M. 2002. Data modelling versus ontology engineering. ***ACM SIGMod Record***, 31(4): 12–17.

Trinkunas J. & Vasilecas, O. 2007. A graph oriented model for ontology transformation into conceptual data model. ***Information Technology and Control***, 36(1).

Truong, C., Oudre, L. & Vayatis, N. 2020. Selective review of offline change point detection methods. ***Signal Processing***, 167: 107299.

Copyright © 2023 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.