Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

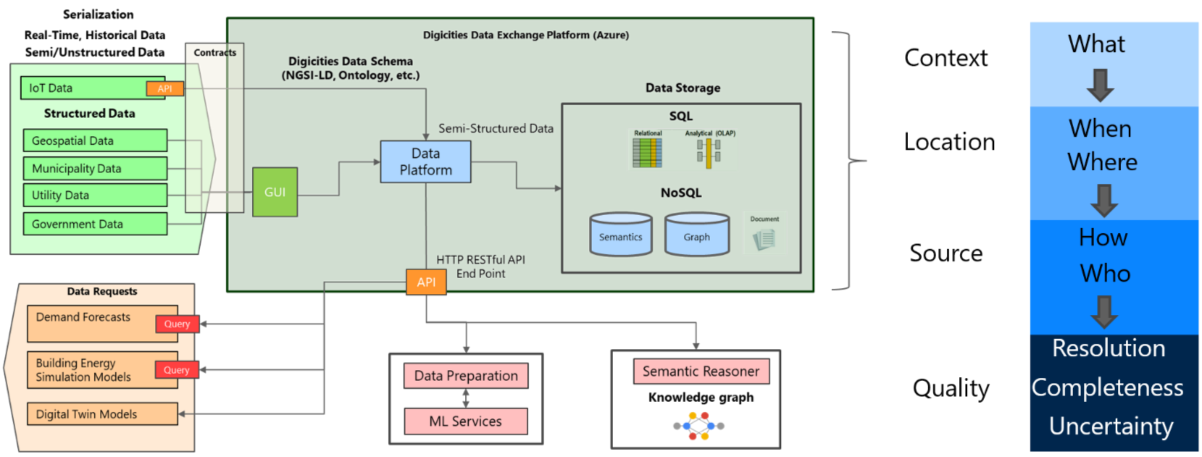
Federal Department of the Environment, Transport, Energy and Communications DETEC

Swiss Federal Office of Energy SFOE
Energy Research and Cleantech Division

Interim report dated 8th May 2023

DIGICITIES

Urban Digital Layers to Support the Energy Transition of Cities



Date: 8th May 2023

Location: Bern

Publisher:
Swiss Federal Office of Energy SFOE
Energy Research and Cleantech
CH-3003 Bern
www.bfe.admin.ch

Subsidy recipients:
Empa
Überlandstrasse 129, 8600 Dübendorf, Switzerland
https://www.empa.ch/

SUPSI
Stabile Le Gerre, 6928 Manno, Switzerland
https://www.supsi.ch/

(L*3) Città di Lugano
Piazza Riforma 1, 6900 Lugano, Switzerland
https://luganolivinglab.ch/

Cividi
Flüelastrasse 10, 8048 Zürich
https://www.cividi.ch/

Authors:
James Allan, Empa, james.allan@empa.ch
Francesca Mangili, SUPSI, francesca.mangili@idsia.ch
Marco Derboni, SUPSI, marco.derboni@idsia.ch

SFOE project coordinators:
Karin Söderström, karin.soederstroem@bfe.admin.ch
Michael Moser, michael.moser@bfe.admin.ch

SFOE contract number: SI/502384.01

Zusammenfassung

Die Energiewende erfordert einen Paradigmenwechsel wie wir Energie bereitstellen, verteilen und nutzen. Die zunehmende digitale Vernetzung und die aufstrebende Industrie 4.0 bedeuten, dass mehr Daten zur Verfügung stehen. Aber es mangelt immer noch an semantischer Interoperabilität zwischen den Datensätzen. Dies schränkt die Entwicklung skalierbarer energieorientierter Anwendungen ein. DIGICITIES' Ziel ist es die Hindernisse für den Zugang und Austausch von Daten zur Entscheidungsfindung bei Versorgungsunternehmen und Gemeinden abzubauen. Es wird eine Datenarchitektur auf der Grundlage strukturierter, miteinander verbundener digitaler Ebenen entwickelt, die für die Berechnung des zukünftigen Energiebedarfs verwendet werden. Die Verarbeitung, Speicherung und Nutzung von Datenquellen werden in einem Living Lab in jedem Partnerland demonstriert. Dem Projektkonsortium gehören Akteure aus allen Stufen der Wertschöpfungskette an. Dieser ganzheitliche Ansatz berücksichtigt die Auswirkungen von technischen Innovationen und regulatorischen Änderungen, um wirkungsvolle Lösungen zu entwickeln, welche die Transformation zu einem Netto-Null-Energiesystem beschleunigt. Dieser Zwischenbericht markiert das Ende des ersten (von drei) Projektjahren. In diesem Jahr haben wir einen Rahmen für die Zusammenarbeit zwischen den Projektpartnern geschaffen, die Beteiligten befragt und den Stand der Technik beim Datenaustausch im Energiesektor untersucht. Die Erhebungen und Überprüfungen wurden genutzt, um einen Prototyp zu erstellen, der zur Spezifizierung der Komponenten der Pilotimplementierung verwendet wurde. Das Projekt tritt nun in die Umsetzungsphase ein.

Résumé

La transition énergétique nécessite un changement de paradigme dans la façon dont nous produisons et utilisons l'énergie. L'interconnectivité croissante et l'essor de l'industrie 4.0 signifient que davantage de données sont disponibles, mais il y a toujours un manque d'interopérabilité sémantique entre les ensembles de données. Cela limite le développement d'applications évolutives orientées vers l'énergie. DIGICITIES vise à surmonter les obstacles à l'accessibilité et à l'échange de données pour la prise de décision à l'échelle des services publics et des municipalités. Une architecture de données sera développée autour de couches numériques structurées et interconnectées qui seront utilisées dans la projection des demandes énergétiques. Un cadre pour le traitement, le stockage et l'utilisation des sources de données sera démontré dans un laboratoire vivant dans chaque pays partenaire. Le consortium du projet comprend des parties prenantes à chaque étape de la chaîne de valeur. Cette approche prend en compte l'impact des avancées techniques et des changements réglementaires pour développer une solution qui accélérera la transition énergétique vers un système à énergie nette zéro. Ce rapport intermédiaire marque la fin de la première année (sur trois) du projet. Au cours de cette année, nous avons établi un cadre de collaboration entre les partenaires du projet, interrogé les parties prenantes et examiné les pratiques d'échange de données les plus récentes dans le secteur de l'énergie. L'enquête et l'examen ont servi à créer un prototype qui a été utilisé pour spécifier les composants de la mise en œuvre pilote. Le projet entre maintenant dans sa phase de mise en œuvre.

Summary

The energy transition requires a paradigm shift in how we generate and use energy. Increasing interconnectivity and the rise of Industry 4.0 means that more data is available but there is still a lack of semantic interoperability between datasets. This restricts the development of scalable energy oriented applications. DIGICITIES aims to overcome the barriers to accessibility and exchange of data for decision-making at a utility and municipality scale. A data architecture will be developed around structured, interconnected digital layers that will be used in the projection of energy demands. A

framework for the processing, storage and use of data sources will be demonstrated in a living lab in each partner country. The project consortium has stakeholders from each stage of the value chain. This approach considers the impact of technical advancements and regulatory changes to develop a solution that will accelerate the energy transition to a net-zero energy system. This interim report marks the end of the first year (out of three) of the project. During this year, we established a collaboration framework between project partners, surveyed stakeholders and reviewed the state-the-art data exchange practices in the energy sector. The surveying and reviewing was used to create a prototype that has been used to specify the components of the pilot implementation. The project is now entering the implementation phase.

Abbreviations

AEM	Azienda Elettrica di Massagno
AIL	Aziende Industriali di Lugano
API	Application programming interface
AUC	Area Under the ROC Curve
BRIDGE	European Commission initiative in smart energy systems research
CIM	Common Information Model
DL	Deep Learning
DSM	Demand Side Management
DTDL	Digital Twin Definition Language
GIS	Geography Information System
GML	Geographic Markup Language
IoT	Internet of Things
kWp	Kilowatt peak
HVAC	Heating Ventilation and Air Conditioning
LIC	Lugaggia Innovation Community
ML	Machine Learning
NGSI-LD	Next Generation Service Interfaces - Linked Data
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
RE	Renewable energy
SDAT	Standardised data exchange recommendation
SES	Società Elettrica Sopracenerina
SFOE	Swiss Federal Office of Energy
SGAM	Smart grid architecture model
SQL	Structured Query Language
UBEM	Urban Building Energy Modelling
URI	Uniform Resource Identifier
XML	eXtensible Markup Language

1 Introduction

1.1 Background information and current situation

1.1.1 Swiss digital strategies

In Switzerland, the federal council has implemented a strategy for a "Digital Switzerland" urging stakeholders from different sectors to implement digital transformation projects (BFS, 2020). 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. Digitalisation is already creating a rapid increase in the volume of data generated and available.

1.1.2 Importance of data exchange for the energy sector in Switzerland

A study into approaches for data exchange in the energy sector commissioned by the Swiss federal office of energy provides a set of recommendations for the sector (Holles et al., 2021). The report estimates that the combined net present value of two data hubs for electricity and gas could reach approximately 1billion CHF/year. The report details the highly decentralised nature of data exchange across the numerous stakeholders of the electrical and gas grids. Digital technologies are expected to help operators of the grid handle the growing complexity and requirements; however, the report mentioned a need to significantly improve access and use of data. To facilitate this, the report investigates options and key features for a data hub to exchange data in the energy sector. Key use cases of such a hub include: the exchange of metering data, representation of flexibility, external access, implementation of change processes and end-user offer management. The report also makes a recommendation on the standardisation of data and the use of application interfaces (APIs) to access the data. The report acknowledges that much of the data used in the operation of the Swiss electricity grid is standardised according to the Standardised Data Exchange Recommendation (SDAT). Beyond Switzerland, a strategic evaluation of the benefits of digital transformation for the energy sector revealed that there are opportunities for digital transformation to optimise energy allocation and scheduling; however, the main barriers are weaknesses, such as resistance to change and security, and threats, such as security risks (Liu and Lu, 2021).

1.1.3 Data availability in Switzerland

Open data initiatives, such as opendata.swiss[1], are managing the influx of data and have the additional challenge of making it publicly available. Each dataset is published along with its metadata, which includes fields such as, language, spatial & temporal coverage, legal statements and contact details. The datasets are prepared according to the organisations mandated to provide data, and, as a result, the data is often heterogeneous in structure, attributes, format and quality. Additional transformation and mapping are often required to link datasets; however, increasing awareness of data standards will improve interoperability. The data also needs to be discoverable and useable for its stakeholders.

The data published on opendata.swiss is from a diverse number of sources including geospatial, health, meteorological and statistical, and many datasets are updated and published periodically. Due to changes in requirements and collection methods, the structure and contents of published datasets may change between updates. For example, certain data fields, might be present in one update and/or absent or relabelled in another. This can lead to issues in interoperability for applications reliant on up-to-date data. Nowadays, real-time integration of data from sensors and meters is becoming increasingly important. The rapidly growing and evolving Internet of Things (IoT) sector, is making

[1] https://opendata.swiss

data collection from such devices possible. IoT is a refers to the interconnection of physical devices, such as sensors and systems to the internet, enabling them to collect, exchange and analyse data to enable new capabilities for applications spanning numerous domains (Gubbi et al., 2013). In the energy sector, IoT can support the implementation of smart grids across all stages of energy supply chain, from generation to consumption (Ghasempour, 2019). IoT devices also provide a data source for load forecasting (Li et al., 2017). Switzerland has the legally binding target of replacing 80% of conventional meters with smart ones by 2027 (Schweizerische Bundesrat, 2008). This will generate a source of live data for the planning and operation of power systems for grid operators and energy planners.

1.1.4 General challenges for data management

The challenge of data exchange in the energy sector is not a problem that is unique to Switzerland. A BRIDGE working group set-up to achieve interoperable and business agnostic data exchange was established to define a common reference architecture for data exchange in the energy sector to support demand-side management (Lambert, Eric et al., 2021). This report included surveys of the adoption of common information models in the energy sector. The report makes reference to the Smart Grid Architecture Model (SGAM) to organise the interaction of processes in the sector (Gottschalk et al., 2017). The report highlights the various bottom-up initiatives to support cross-sector and cross-border data exchange (GAIA-X, FIWARE, Data Bridge Alliance, IDSA, Open DEI). The report recommends developing cross-sector data models, to facilitate data exchange between sectors. The report recommends a data format agnostic approach to cross-sector data exchange and universal data applications that can be implemented in different domains. BRIDGE Projects are increasingly using business process agnostic platforms such as Ecco SP, Estfeed, IEGSA, Atos FUSE, Enterprise Service Bus and Cloudera. The CIM models are standards that are unique to the energy industry. They are not open-access standards, which makes cross-sector mapping challenging. The information listed in SGAM are not limited to CIM but also include: COSE, IEC 61850, SAREF, CIM+, NGSI-LD, FIWARE and SAREFwater (Gottschalk et al., 2017).

Data streams between system operators, consumers, producers and other energy stakeholders are becoming larger and more diverse, with new requirements on data exchange being introduced by the efficient operation of buildings and the integration of new technologies into the energy system. Semantic interoperability between applications like building energy control, grid operation and energy system modelling is therefore fundamental to the digitalization of the energy infrastructure (Pritoni et al., 2021).

1.1.5 Data quality challenges

Data Cleaning is the process of detecting errors in a dataset and repairing them to improve the quality of the data. This includes qualitative data cleaning which uses mainly integrity constraints, rules, or patterns to detect errors and quantitative approaches which are mainly based on statistical methods (Thirumuruganathan et al., 2020). Specific sub-problems tackled by data cleaning include:

- **Missing data,** due to device malfunction, outages of the wireless network, consumers' behaviours, etc.
- **Data dependencies**, e.g., between similar meteorological variables or highly dependent smart meter measurements, and data redundancies, e.g., repeated buildings.
- **Outlier detection**, identifies anomalous data that does not match a group of values, either syntactically, semantically, or statistically. For example, jumps in the dates or unexpected zeros the consumption records may indicate anomalous records.
- **Non-homogeneous time scales**, between time series collected from different buildings or systems, ranging from 10/15 minutes for smart meter to one month for more traditional

acquisition of consumption data. This is particularly important if high temporal resolution forecasts are needed also for buildings not equipped with smart meters.

- **Inconsistent data**, for instance in hierarchical data the values on a level of the hierarchy do not match the sum of the observation in the lower levels.

1.1.6 Architectures to support energy data

In the SGAM, information models are an integral part of the smart grid architecture. Ontologies can also be used to integrate energy management data on the building and district level (Hippolyte et al., 2018). The limitation of this approach was that the mapping between measured data and ontology needed to be done manually, which is a time-consuming task.

1.1.7 Importance of load forecasting

Load forecasting is the process of predicting future energy demand using historical data and its influencing factors. Load forecasting can help achieve the following:

- **Optimal resource allocation and investment planning:** The decision making process in the energy industry is reliant of 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 (Hong and Fan, 2016). The use of demand forecasting can also be used as inputs to energy system optimisation models to support the decision makers (Scheller and Bruckner, 2019).
- **Integration of renewable energy resources:** Incorporating the integration of distributed energy resources can lead to a more holistic modelling of smart grids as it enables an understanding of how the predicted loads can be met with intermittent generation from renewables (Habbak et al., 2023).

1.1.8 Load forecasting with physics-based simulation

Physics-based models simulate the energy generation or the demand of systems connected to the energy grid. Physical models use data as parameters and variables to model the thermodynamic process that occur in a system.

Physical models of buildings, model the energy flows between building components in response to environmental factors such as weather conditions, to estimate the resulting energy demand. Building energy simulation is commonly used to evaluate the environmental performance of a building during the design phase, where there is a need for models that achieve an accurate representation of the real-world building operation (Coakley et al., 2014). 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 (Hong et al., 2020). UBEM models use data obtained from city datasets such as geometries and building statistics. UBEM models require significantly less data than detailed models and parameters are often assigned using archetypal approaches (Wang et al., 2018).

Thermodynamic models of varying complexity are established for all forms of renewable technology. For example, the performance of solar photovoltaics, solar thermal and wind can be easily calculated using simple models that capture the behaviour of such systems in response to irradiance and wind with reasonable accuracy (Duffie et al., 2020). The inclusion of these models in open-source engineering libraries of technologies and components enables them to be integrated and simulated as part of complex systems (Wetter et al., 2014).

1.1.9 Machine learning to support load forecasting

A plethora of machine learning approaches have been proposed in the recent years in response to the increased availability of data, and in particular smart grid monitoring data. These methods have been shown to achieve a high accuracy, however they have seldom be studied in the context of a real or realistic framework. High performance in terms of accuracy or AUC (Area under the ROC Curve) on benchmark data does not necessarily results in effectiveness in real situations. In fact, some requirements may be overlooked in the context of ML development. For instance, awareness about the uncertainty of the forecasts produced may be useful when it comes to decision making and control. However, lots of the algorithms preferred for their superior performance, such as neural networks and tree-based gradient boosting frameworks may fail providing reliable uncertainty estimates. On the other hand, so-called statistical approaches (e.g., auto-regressive moving average) provide uncertainty estimates but cannot capture complicated non-linear relationship between the output and the input predictors, and thus cannot achieve the same accuracy as black-box ML models. It must be also considered that, while ML methods are often validated on clean benchmark data, real data are less refined and include outliers, missing values and different resolutions.

(Ferrero Bermejo et al., 2019) investigated different RE sources, such as solar, wind, and hydro, and empirically proved the efficiency of the artificial neural network for power prediction. Similarly, ML applications and taxonomy have been briefly studied (Mosavi et al., 2019) for energy systems, where, based on experiments, the authors analyzed that the hybrid models showed accurate prediction scores. Furthermore, numerous forecasting algorithms have been examined from diverse viewpoints including energy policy, economy, battery storage capacity, and power generation in RE sources (Ahmed and Khalid, 2019). Later, the researcher's attention was diverted to the deep learning (DL) for its remarkable performance in prediction tasks because of its strong capabilities in recognizing the primary nonlinear characteristics rather than employing handcrafted features (Zhu et al., 2022).

Forecast reconciliation is a process by which forecasts generated independently for a collection of linearly related time series are combined according to such constraints to obtain more accurate prediction than the original ones for each time-series (Meira et al., 2023). The underpinning principle is to exploit the combination of multiple forecasts at different levels of granularity to reduce the uncertainty of each individual forecast. (Hollyman et al., 2021) reviews the recent literature about reconciliation and shows how even simple approaches to reconciliation may provide significant improvements in forecasting accuracy, especially at the bottom levels of the hierarchy. In the energy field, cross-sectional reconciliation is for instance used to exploit hierarchies imposed by the network structure. The most recent approaches to forecast reconciliation, which exploit at the same time temporal and cross-sectional hierarchies(Spiliotis et al., 2020) (Di Fonzo and Girolimetto, 2023), are promising for application to energy time series forecasts at different spatial and temporal levels of granularity.

Transfer learning is an effective method to solve the problem of modelling with small sample. Its main idea is to transfer the rules or knowledge learned from source prediction tasks that possess sufficient data to the target prediction task with insufficient data, with which the learning of target task could be facilitated under the condition of limited data (Lu et al., 2021).

1.2 Purpose of the project

Digicities will develop, implement and demonstrate a data exchange platform to meet a set of use cases defined through conducting surveys with the data contributors and the need-owners of the exchange platform. The data exchange platform will act on several of the recommendations for a data hub in (Holles et al., 2021): the simplification of data structures, use cases for data access and innovation, extension of the existing data architecture to consider new sources of data. The exchange platform will demonstrate how cross-sector data can be mapped using standardised models to predict the energy demand at a specified output level. This is also a challenge identified in the BRIDGE report (Lambert, Eric et al., 2021). The resulting architecture and data models will be mapped against the SGAM model (Gottschalk et al., 2017). The platform will accommodate the supply of parameters for both UBEM and ML studies. The critical functionality of the platform is to provide structured semantic information relevant to the use case. Although novel services and technologies will be applied in this project, the process will be documented and technology agnostic so approaches could be implemented in other regions and use cases. The datasets published by opendata.swiss will be integrated as a core component of the platform and this project serve as a demonstration of how the datasets can be applied for energy demand forecasting, either using ML models or UBEM. This project is aligned with the objective of increased renewables in the energy strategy (SFOE, 2022), where models will be used to evaluate the impact of intermittent generation from renewables and plan according; and the digital strategy for Switzerland (BFS, 2020), where the project will contribute to data transformation activities.

1.3 Objectives

This project objectives and objectives were conceived at the beginning of the project and are listed below:

1.3.1 Objectives

- Provide semantic data layers to support existing digital platforms and services
- Enable the projection of energy demands to support decision-making
- Implement an architecture that helps organisations achieve secure data exchange
- Use the data architecture to evaluate renewable energy integration and energy efficiency measures
- Investigate new business opportunities to sustain the platform

1.3.2 Outputs

- Generation of additional features to train machine learning models
- Guidelines exchanging and processing energy data for the demonstration use cases
- Demonstrated pipelines focused on improving data quality
- Digitalisation pathways leading to the increases in energy efficient and renewable technologies

In addition to the above objectives the project seeks answers to the following research questions, as defined in the proposal:

Technology: Implement a data infrastructure to enable utilities and municipalities to make better use of their data for short-term and long-term energy planning.

> **RQ T.1** What are the under-utilized data resources that are available to utilities and municipalities to make more accurate demand projections and better decision making?
>
> **RQ T.2** What are the key semantic attributes of the data that are needed to facilitate data exchange across the digital value chain?
>
> **RQ T.3** How can machine-learning techniques be used to address challenges of data incompleteness and inaccuracy?
>
> **RQ T.4** How can federated queries be applied to multiple heterogeneous data sources of urban data and used to project the energy demand of buildings?
>
> **RQ T.5** What are the key components of the infrastructure to ensure security and sovereignty of data exchange between the necessary stakeholders of the value chain?

RQ T.6 How can data be used for short and long-term decision-making regarding the integration of intermittent renewables, storage technologies and building retrofitting?

RQ T.7 How can multi-source live-data be seamlessly integrated into a digital twin and used for demand balancing?

Commercial: Address the challenges of data integration in the energy planning process.

RQ C.1 What is the business case in Switzerland for adopting a structured data architecture for the exchange of urban energy?

RQ C.2 What next steps are needed to make the process transferable and scalable to other municipalities and districts in Switzerland?

Stakeholder: Develop and assess the feasibility of business opportunities for urban energy data.

RQ S.1 What are the key drivers of each stakeholder group to participate in the exchange of energy data for urban modelling?

RQ S.2 What are the key features that need to be integrated into a national dissemination plan to maximize consumer acceptance and awareness?

RQ S.3 What is the impact on existing energy management tools used for decision making by municipalities and utilities?

RQ S.4 What are the main benefits for citizens and what incentives are needed to engage them to strengthen the data exchange architecture e.g. crowd sourcing initiatives?

RQ S.5 What are the impacts of the data architecture and newly available data resources on existing practices?

RQ S.6 How suitable is the data for academic research and are there any additional features that must be considered?

2 Description of facility

This pilot and demonstration project will be implemented using data generated in three different sites across Switzerland.

2.1 Lugaggia Innovation Community

The Lugaggia Innovation Community (LIC) pilot. The project was launched in March 2019 by a consortium in the Southern Switzerland, to promote an experimental self-consumption energy community called Lugaggia Innovation Community, which connects the Lugaggia kindergarten with 18 neighbouring houses and 5 photovoltaic plants (with a total power of 90 kWp). The objective of the project was to study the application of active DSM on a set of electrical loads spread in a residential area with the presence of renewable sources as well as energy storage systems (Salani et al., 2020). This was achieved by a centralised energy management platform using a smart meter infrastructure for sensing and actuation and a decentralised control approach managing the use of energy by the community devices through computing and controlling units, connected to the smart meters.

2.2 Lugano Living Lab

Lugano Living Lab is a platform that aims to encourage innovation in Lugano, with the aim of improving the quality of life of citizens and the attractiveness of the region, through an innovation oriented approach based on dialogue, co-creation and collaboration of all the existing forces active in the area. Lugano Living Lab works directly with the City of Lugano municipality and its main energy utility company, AIL.

2.3 NEST Building

The NEST is an experimental building comprised of interconnected modular units. Energy is produced on-site using integrated renewable technologies or supplied through energy networks. Energy is supplied to the units through a multi-energy hub, which is comprised of a microgrid and three thermal networks. The isolation and control of different parts of the network allows technologies and control strategies to be evaluated. Approximately 10,000 measurements are recorded every minute and stored in a time series database and can be queried through a REST API. These measurements include the heating and cooling demands from the thermal network, as well as HVAC, lighting, and equipment electricity consumption. Metadata about each data point is stored in a SQL database.

3 Activities and results

3.1 Use case surveys

The purpose of profiling the inputs and outputs of our digital services is to understand the types of data they need and the value they can generate on it. This will enable a common understanding of how to prepare and exchange data for the data generators and the third parties. This will also help us identify conflicts of interest. It is also possible for the need-owners to work on the data themselves but this is often outside the scope of their business activities.

3.1.1 Survey of present data usage of need owners

The need owners of this project are energy utilities. Three energy utilities provide energy to the region of this study and participated in the Digicities kick-off meeting (AEM, AIL, SES). The following questions were asked about the present operation of the company:

- *What is your company's offering?*
- *How important is INTERNAL data in your actual/future business.*
- *How important is EXTERNAL data in your actual/future business.*
- *What INTERNAL data are you currently using?*
- *What EXTERNAL data are you currently using?*
- *How are you obtaining INTERNAL data?*
- *How are you obtaining EXTERNAL data?*
- *What are the pains obtaining data?*
- *What are the pains dealing with data?*
- *What are the main uses you make of data?*
- *How do you deal with low quality data (wrong/missing samples)?*

3.1.2 Survey of present future usage of need owners

The utilities were then asked the following questions about their intended future use of data:

- *What data are you missing?*

- *What are the hurdles obtaining that data?*
- *What are you doing to fill the gaps?*
- *Are you generating proprietary data?*
- *Would you be willing to sell your proprietary data on a data exchange?*
- *Would you need raw data or pre-defined Insights? Which ones?*
- *What are the NEW uses you would make of new data / insights?*
- *Would certification of a data exchange be valuable in your business?*
- *What would be a "dream tool" based on data that would boost your business?*

3.1.3 Survey of digital services

Digital services are offered by organisations that use data to generate energy insight for their clients. Several organisations joined the Digicities knowledge community during the project kick-off. The survey of digital services was carried out to identify the main datasets required to generate insight for the clients. The following questions were asked:

- *What is your company offering?*
- *What data are you currently using?*
- *How are you obtaining data?*
- *What are the pains?*
- *What data are you missing?*
- *What are the hurdles obtaining the data?*
- *What are you doing to fill the gap?*
- *Do you generate proprietary data?*
- *Would you be willing to sell your proprietary data?*

3.1.4 Use case definitions

The responses to the survey were analysed during a workshop and the following use cases were defined:

Utility use case 1

We need to know what a district's energy consumption and production patterns will be in 2030 to plan and build a dependable and competitive network. Our infrastructure department needs to optimize its network to guarantee stable and secure operation in the future. Our operations department needs to predict the consumption over the short-term (24 hours) and medium-term (3 months) horizons to decide whether to produce or import energy. This requires continual monitoring of the grid operation and its available capacity. Long-term (1-year plus) projections are needed to make strategic decisions. Our operations also need to be reactive to local events at a time resolution of 15min so that the economic and environmental performance of the grid can be maintained.

Utility use case 2

We would like to know how our grid will function in the future. We would like to know what actions we should take now to develop our grid to ensure reliable and sustainable operation in the future. We would like to access the best available information to inform us how our grid will respond to different scenarios and external factors such as resource availability and future energy consumption patterns. We believe a digital platform with continually updated information on the technical advancements in energy consumption, production, storage and distribution would facilitate decision-making for the development and operation of our future grid.

NEST Demonstrator Use Case 1

An exemplary use case using the NEST demonstrator building of smart meter interaction at utility and building level. In the Digicities project, the NEST use case would primarily act as a data contributor to demonstrate how the different data scales can be handled between the system and the building scale.

Energy Simulation Use Case 1

Both use cases aim to explore the building energy consumption patterns and the interaction with the neighborhood energy systems, such as the public electric grid and district heating network. This aims to improve and maintain the stability and reliability of energy networks. In this regard, a data architecture, compliant with current and upcoming standards that preserve sovereignty, security, and privacy, needs to be implemented. Furthermore, we will gain more insights into how the use of open and standardized APIs will maximize data usability and contribute to supporting achieving the city’s energy and climate goals.

3.2 Design of the conceptual architecture

The requirement of the conceptual architecture was designed according to the results from the stakeholder survey workshop. The process is shown in Figure 1.

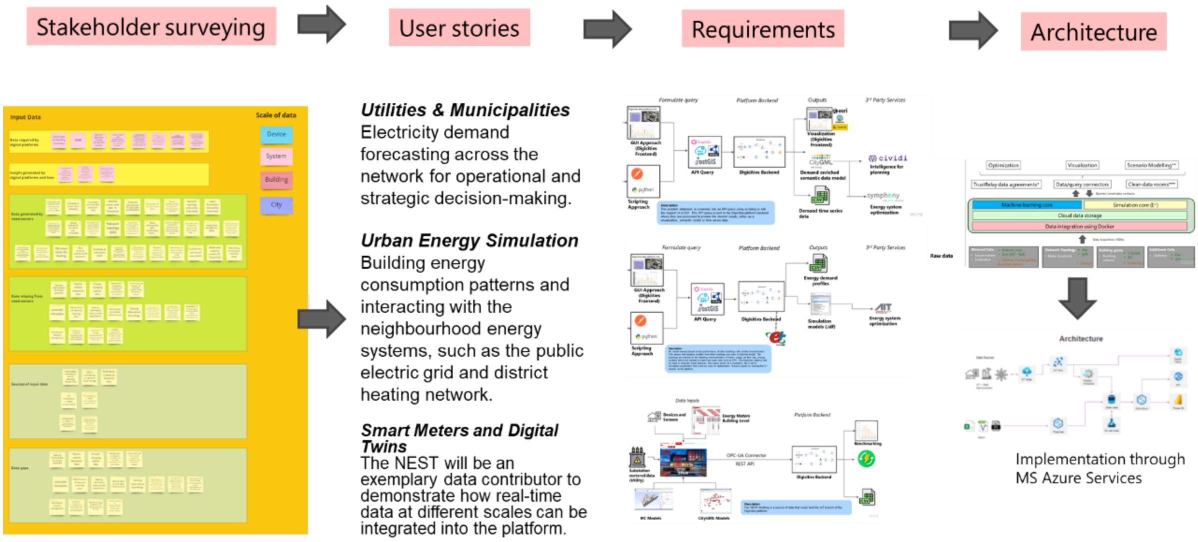


Figure 1: The process to design the conceptual architecture of the Digicities platform. The results of a stakeholder survey were carried out to determine the user stories. This was used to generate a set of requirements from which a conceptual architecture was prepared.

Once the architecture was mapped out it was sent to our implementation partner (Microsoft) to identify the cloud services that could achieve the requirements.

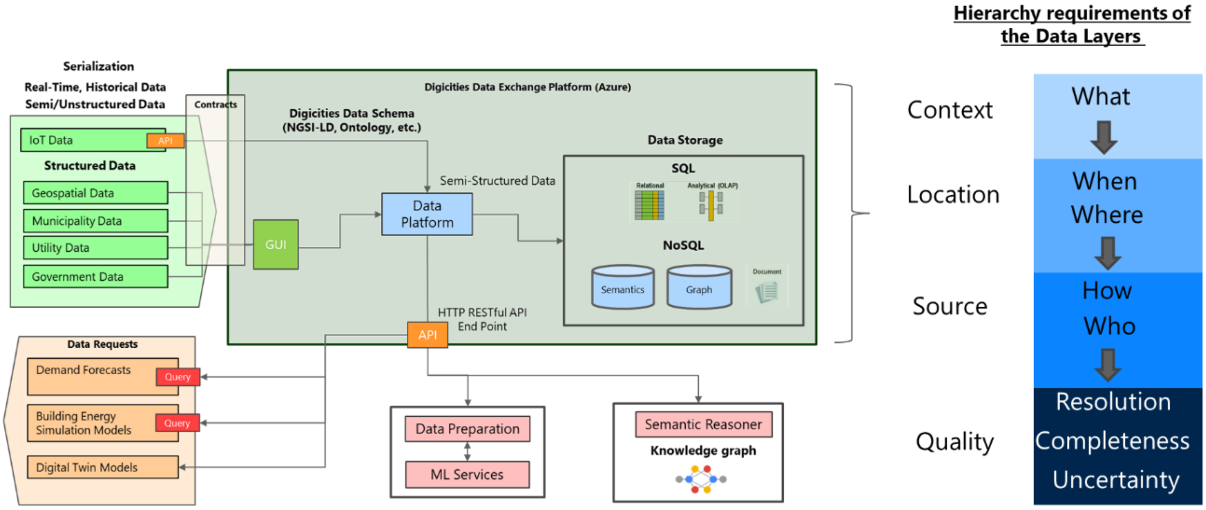


Figure 2: A closer look at the specific requirements in relation to data contributions and data requests. The data requests and contributions are handled by the Digicities data platform. The platform has a data storage ecosystem that enables data to be stored optimally according to the data format and type, this can be separated into SQL and NoSQL databases. The data is made available through an API that enables federated queries across the data stored on the platform. Data preparation and ML services are designed to work directly with the platform. The use of standard vocabularies to organize data enables semantic reasoning. All data stored on the platform must adhere to the hierarchy of semantic requirements shown on the right. This is to assist querying and enable the development of scalable and transferable algorithms.

3.3 Review and categorization of available data

3.3.1 Data interactions

The platform will have two forms of interaction: contributions and requests. The business case for closing the loop by providing services enable by data enhanced by the Digicities platform, will be evaluated in this project. An overview of the interactions is shown in Figure 3.

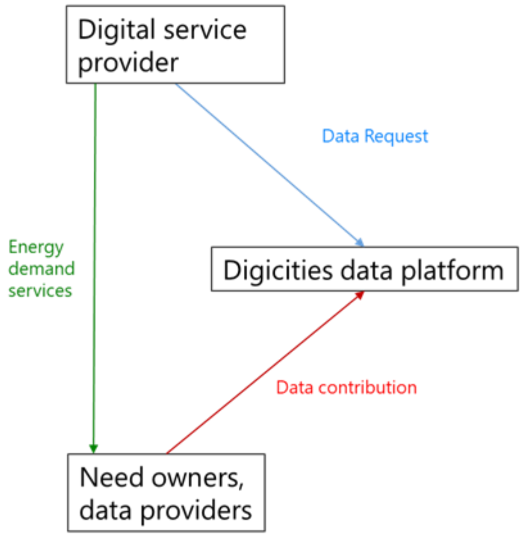


Figure 3: Data interactions foreseen in the Digicities project

3.3.2 Data governance

A governance framework defines the rules of working with the data integrated and used by the platform. We have categorised the data received in this project by access and usage restrictions. These options will be presented to data contributors or when establishing data agreements. The different categories of data access are:

- Open data: The data is published online and can be accessed by anyone.
- Commercial data: The data is sold commercially.
- Privately shared data: The data is shared by an organisation for a specific purpose.

Each type of data access will come with its own terms and conditions for data usage. The following set of usage restrictions are proposed to specify the access rights.

General use: The use of the data is limited to a specific purpose e.g. a research project or non-commercial usage

- Data linking: The data cannot be linked with other datasets.
- Personal and sensitive data: Data must not be shared with third parties.
- Reporting restrictions: Data must be reported at a level of aggregation so that the underlying data cannot be identified.

3.3.3 Data domains

The features of each dataset are also classified by domain. Features in each dataset were assigned to the following domains:

- Buildings
- GIS Boundaries
- Energy and time series
- IoT and real time

3.3.4 Semantic requirements

The integration of data according to common information models or ontologies enables interoperability. The features in each dataset are evaluated according to several requirements that determine context, location, source and quality. The hierarchy of the semantic requirements of the data layers are shown in Figure 4.

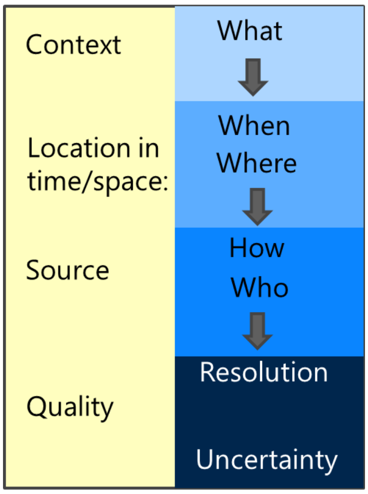


Figure 4: Semantic requirements of the data layers in the platform

3.4 Review of available datasets

The data that will be integrated into the platform have been classified according to the above considerations. An overview of the datasets identified during the first year of the project are shown in Table 1.

Table 1: Summary of Swiss datasets

Dataset	Features	Role	Restrictions
GIS boundaries	GIS	Core	Open
Open Street Map	Building	Core	Open
Swiss Buildings 3D	Building	Core	Open
Register of Buildings and Dwellings (GWR)	Building, Energy	Core	Open
Weather	Weather	Core	Open
Network topology	Energy	Contribution	Restricted
Municipality data	Building	Contribution	Restricted
Smart meter	Energy	Contribution	Restricted

3.5 Review of data vocabularies

3.5.1 Geospatial vocabularies

Geospatial data contains geographic or spatial information. Geographic Information System (GIS) is the software system used to work with geospatial data. GIS software enables layer-based analysis of vectors (e.g. points, lines, polygons) and raster data (e.g. satellite images, tiles). Geospatially resolved data is increasingly available and there are a large number of applications to manage and visualise the data. They are an essential component of data platforms for cities. A review of the different technologies for storing and analysing geospatial data found that document databases are the best platform for geospatial data processing; whereas graph databases and key-value databases are limited in both data structures and the processing options (Guo and Onstein, 2020).

GeoNames Ontology

Type: Web Ontology Language

Physical Scale: Regional

Temporal Scale: NA

Geonames is a database that contains place names of for all regions across the world, in multiple languages. It is a collaborative effort and contains contributions from many organisations in each country. The GeoNames Ontology is designed to enable the use of geospatial information on the semantic web2.

GeoSPARQL

Type: Web Ontology Language

Physical Scale: Can represent any scale of geospatial object

Temporal Scale: NA

GeoSPARQL is a web ontology language for representing and querying geospatial data using semantic web technologies. It is developed and maintained by the Open Geospatial Consortium (OGC). It was created to enable geospatial data interchange with efficient geospatial queries (Battle and Kolas, 2011). GeoSPARQL only defines two subclasses of Spatial Object: Features and Geometries, which can both belong to spatial collections.

3.5.2 Formalising spatial scales for urban modelling

The spatial scales are mostly administrative constructs that are specific to a region or country. For example, there is no universal definition of a postcode, district or canton. These will mostly vary from country to country. A spatiotemporal ontology for administrative units of Switzerland (SONADUS) was created (Gantner, 2011); however, the ontology became large and complex due to adherence to rules in the upper ontology (Gantner et al., 2013). One of the largest resources of linked data on the internet is DBPedia, which also contains information on the Swiss municipality borders. A study showed that querying Swiss regions with the GeoNames terms and creating manual links to DBpedia, returned acceptable performance (Grütter et al., 2017).

3.5.3 Hexagonal hierarchical geospatial indexing system (H3)

H3 is a hierarchical geospatial indexing system for geographic data. Indexed data can be joined across different datasets and aggregated at different levels of precision. The H3 index is an unsigned 64-bit integer representing any H3 object.

[2] https://www.geonames.org/ontology/documentation.html

3.5.4 Building Vocabularies

Brick Ontology

Type: Web Ontology Language

Physical Scale: Device and building scale

Temporal Scale: Relationships connect external time series database

The objective of the Brick ontology is to capture the concepts and relationships necessary to operate a BMS across a heterogeneous set of buildings (Balaji et al., 2016).

Real Estate Core

Type: Web Ontology Language

Physical Scale: Device and building scale

Temporal Scale: Relationships connect external time series database

The Real Estate Core (REC) Ontology is a Web Ontology Language that was developed to support energy usage analysis/optimisation and presence analysis (Hammar et al., 2019). It is designed to accommodate all of the data requirements of real estate management. The REC is comprised of two base models that can be extracted. The REC has the following modules:

- Metadata
- Core
- Agents
- Building
- Device
- Lease

GML and CityGML

Type: Conceptual data model

Physical Scale: City infrastructure

Temporal Scale: Relationships connect external time series database

The Geography Markup Language (GML) is an XML-based language for describing geographical features. GML is an open standard maintained by the OGC. CityGML extend the concepts of GML for the representation and exchange of 3D city models. The CityGML is a conceptual data model used to represent virtual 3D city and landscape models (H. Kolbe et al., 2021). The CityGML conceptual data model is comprised of modules that represent different elements of a city. The CityGML conceptual model is extendable through application domain extensions (Biljecki et al., 2018).

An OWL ontology for the CityGML 2.0 schema (University of Geneva, 2023) was unable to represent all of the classes for a test city dataset (Charlottenburg-Wilmersdorf district of Berlin); this lead to a proposed extension of the ontology to represent the required classes in the test dataset (Chadzynski et al., 2021). The authors named the ontology "OntoCityGML" and concluded it could act as the schema to serve a semantic twin to the 3DCityDatabase software; however, at the time of writing, the ontology has not published. 3DCityDB is a geo database to store represent and manage 3D city models using a relational database (Kolbe et al., 2013). At the time of writing, 3DCityDB is compatible with the CityGML 2.0 datasets.

The latest version of CityGML, CityGML 3.0, can now be qualified by a relation type identifiable using URI (e.g. using the sameAs relation from OWL), which allows for mapping to RDF triples (Kutzner et al., 2020).

Applications and technologies work with different modules of the CityGML conceptual model by defining an implementation specification. This contains the results of a set of tests to demonstrate conformance in representing the modules. The only mandatory tests for any application or technology is the CityGML core. CityGML 3.0 contains a new Dynamizer module, which enables: data structures to represent time series data, overwriting of static attributes and explicit linking of sensor and observation data. This module aims to help the integration of IoT devices and the time-series data generated by scenario modelling. The mapping between GeoSPARQL and the core GML concepts has been shown to be feasible (Qiu et al., 2015). The standard CityGML model can be extended using application domain extensions. The most relevant for this project is the EnergyADE (Agugiaro et al., 2018).

IFC

Type: Standard data model

Physical Scale: Device to building

Temporal Scale: NA

Industry Foundation Classes (IFC) is a data model used in Building Information Modeling (BIM) to represent and exchange information about buildings and construction projects. It is an open and neutral standard developed by buildingSMART. IFC is the main reference standard for exchanging BIM models, however, their complexity results in inconsistencies that has limited their use for applications such as energy performance simulation (Elagiry et al., 2020).

BOT

Type: Standard data model for building topologies

Physical Scale: Building

Temporal Scale: NA

The Building Topology Ontology (BOT) is a standardised way of representing the topological components of a building. It is a simplified approach with a limited number of classes that is designed to achieve interoperability between more complex data representations of buildings such as IFC.

gbXML

Type: Standard data model

Physical Scale: system to building

Temporal Scale: sub hourly

Green building XML is an open schema for exchanging building information between building design and energy analysis software tools. It is primarily designed to store information on the Building Energy Model (BEM), which requires specific information on the zones and heating system. As this information is not required for the IFC, there and there are interoperability issues with the IFC BIM format and the gbXML BEM model (Bastos Porsani et al., 2021).

3.5.5 Energy and time series

ASHRAE 223p

Type: Data standard / Web Ontology Language

Physical Scale: Device and building scale

Temporal Scale: Relationships connect external time series database

ASHRAE 223p aims to provide a data standard or tagging dictionary that allows interoperability between building data. This explicitly includes other building data schemas like BRICK, Project Haystack, and most importantly BACnet, which are directly involved in the development of the ASHRAE 223p dictionary. While the development does not seem to be finished, it is aimed to be adopted as ISO standard.

DTDL

Type: Modelling language

Physical Scale: Device and building scale

Temporal Scale: Relationships connect external time series database

The DTDL is comprised of six metamodel classes that are used to describe the behaviour of all digital twins (Azure, 2022). These metamodel classes are Interface, Telemetry, Property, Command, Relationship, and Component. DTDL is implemented in JSON-LD. The Azure Digital Twin Platform incorporated the DTDL.

FIWARE

Type: Standardised data models

Physical Scale: Device and building scale

Temporal Scale: Relationships connect external time series database

FIWARE data models are a collection of standardised data models designed to facilitate the development and interoperability of complex applications that require data from multiple cross-cutting domains, such as energy and smart cities. FIWARE models are defined using JSON-LD, which makes them compatible with linked data principles. FIWARE data models are designed to support IoT interoperability and data exchange between data providers and data consumers using data brokers (Cirillo et al., 2019).

CIM

Type: Standard for data exchange

Physical Scale: Device to system

Temporal Scale: Sub-hourly

The Common Information Model (CIM) is an open data model standard specifically for the electric utility industry to exchange data between different applications and systems. The objective is to improve the interoperability of smart grids; however, the application of CIM is hampered by several issues such as an inability to represent all business requirements, harmonization and validation (Kim et al., 2020). The uptake and application of CIM was the focus of the surveys carried out by (Lambert, Eric et al., 2021).

SAREF

Type: Standardised data model

Physical Scale: Device to system

Temporal Scale: Sub-hourly

Smart Appliances REFerence (SAREF) ontology is a standardized data model developed by the European Telecommunications Standards Institute (ETSI) to enable interoperability among smart appliances and services in the Internet of Things (IoT) domain. SAREF is a data model that focuses on the functionality of smart objects in the building domain. The aim is to enable interoperability in complex systems of connected, heterogeneous devices (Daniele et al., 2015).

3.5.6 Multi-domain models and dataspaces

GAIA-X Dataspaces

Type: Virtual environments for data economy

Physical Scale: Device to city / multi-domain

Temporal Scale: NA

GAIA-X is a European initiative that aims to create a federated, secure and trustworthy infrastructure. Dataspaces are virtual environments designed to handle the requirements of data exchange. GAIA-X is at the heart of the coordination of the European Data Act and Data Governance Act. The aim is to provide use cases and technical architectures for European common dataspace (Braud et al., 2021).

3.5.7 Summary of the data models and platforms considered

The Digicities platform will map the available data and contributions from our stakeholders to the most suitable data vocabulary to represent the information. The conformance to these data models will be documented on the platform. This will enable the discovery and exploration of the data managed by the platform. Figure 5 shows an example of the data models reviewed during this project.

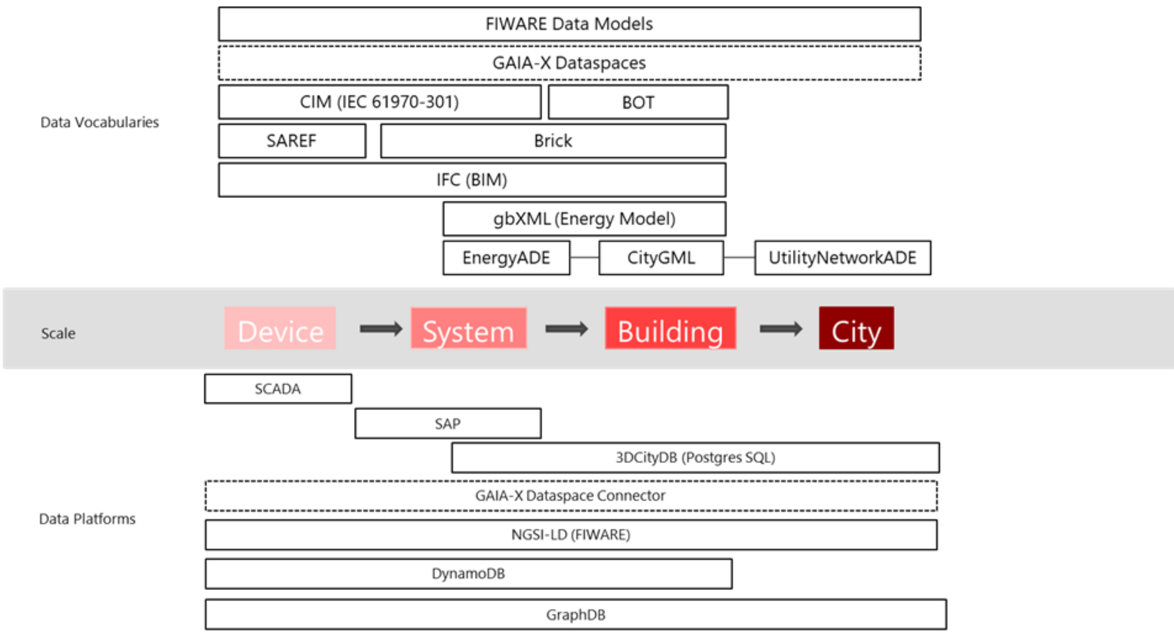


Figure 5: A summary of some of the reviewed data vocabularies and data platforms. The scale indicated the physical scale of the data they represent.

3.6 Project Impact Evaluation

A measurement concept and strategy have been devised to continually monitor and evaluate the progress and impact of the project to ensure it stays on track towards its objectives. The monitoring strategy evaluates each stakeholder and pilot region according to a set of KPIs covering the following impact categories:

- Data access
- Decision making
- Impact on the energy sector

A living report containing details of the pilot regions, planned activities and findings will be updated and submitted to the SFOE bi-annually.

4 Evaluation of results to date

A prototype semantic infrastructure has been created to assist with communicating the necessary components and platform specification to be implemented into the pilot with our implementation partner Microsoft. Data agreements are still in the progress of being finalised with the data providers of our use cases. In the absence of this data we have developed a workflow using simulated data that demonstrates the necessary components that need to be present in the implementation of the pilot platform. The workflow is shown in the following diagram:

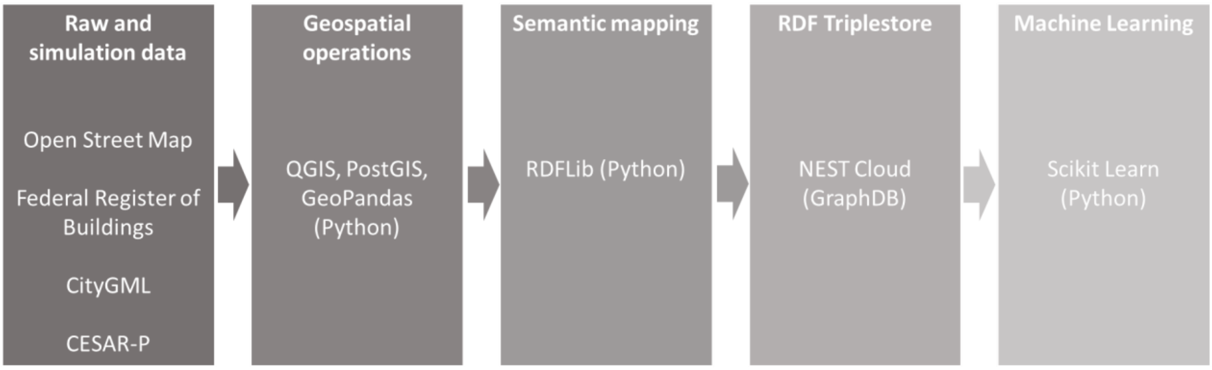


Figure 6: Workflow showing the stages of the prototype ahead of the pilot implementation phase

4.1 Raw and simulation data

Data from the GWR[3] was spatially joined to building polygons in Open Street Map and an energy simulation was carried out using CESAR-P (Orehounig et al., 2022).

4.2 Geospatial operations

The data was spatially mapped to two levels of H3 geospatial cell using spatial joins using the QGIS software package. The process of geospatial mapping is shown in Figure 7.

[3] https://www.housing-stat.ch/de/index.html

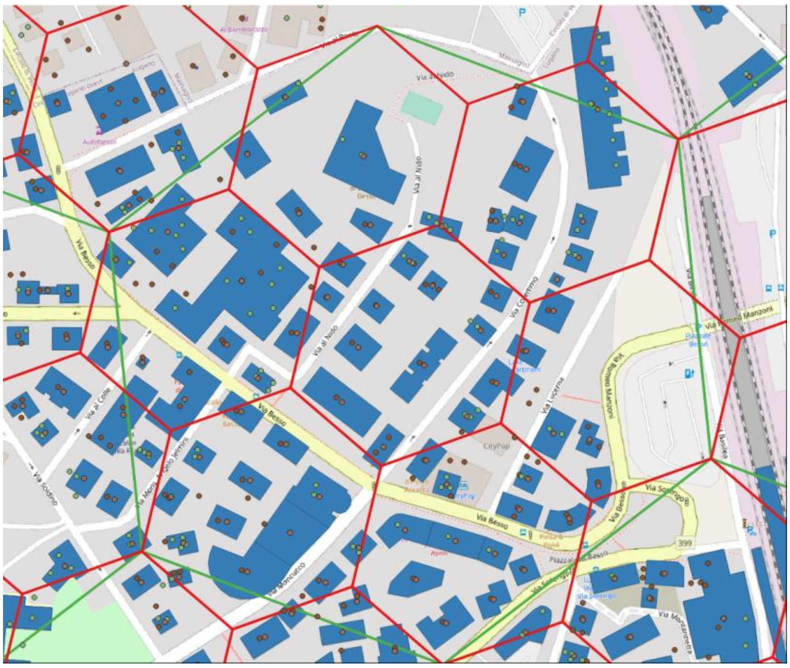


Figure 7: Geospatial mapping of data points to cell grids in QGIS. The red hexagons are resolution 10 of H3 indexing system. The parent hexagon is resolution 9. All hexagons are uniquely identifiable by their ID.

4.3 Semantic mapping

The ontologies used to achieve the semantic requirements for each piece of data are shown in Figure 8. The "What" requirement requires context about the physical object. In the prototype, the UES ontology is used to describe the type of simulation and its parameters (Allan et al., 2021). The Data Catalog ontology[4] (DCAT) is a widely used ontology to facilitate interoperability between data catalogs published on the Web. This is implemented as a core ontology to organise the raw datasets and their distributions. The metadata fields are closely matched to the data fields used to manage open government on opendata.swiss. It will also support hybrid approaches, which incorporate a mix of methodology from each discipline. The PROV-O also enables tracking of changes to datasets and the creation derived information. In the case of UBEM, there are a range of modelling approaches that can generate demands for an entity on the Digicities platform.

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

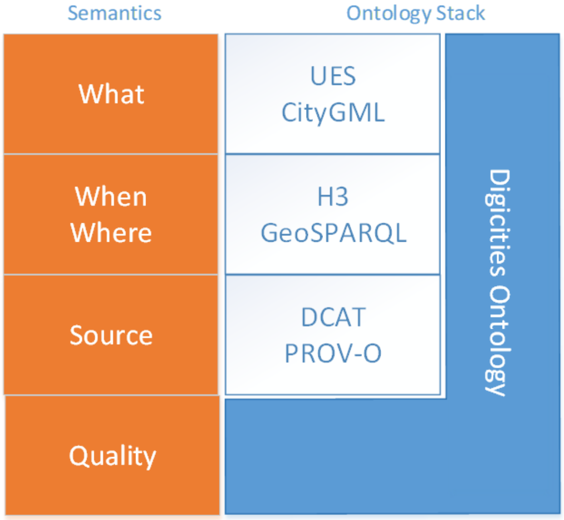


Figure 8: The ontologies used in the mapping of the prototype data to meet the sematic requirements

The provenance ontology (PROV-O)[5] enables the representation, exchange, and integration of provenance information generated in different systems and under different contexts. In the prototype implementation, the classes in PROV-O are used in the creation of a virtual building, which are linked to a set of parameters required for a specific purpose. In the case of the prototype, the virtual entity is linked to input parameters for the simulation. In the future, these virtual entities could be any form of any real-world entity on the platform e.g. building, geospatial cell, administrative region etc. And an unlimited number of virtual entities can be connected to the records of real-world entities. This enables a traceable and repeatable process, which can be used to compare the performance of different modelling techniques e.g. machine learning vs physical models. The application of DCAT and PROV-O achieve the "Source" requirements of the Digicities prototype and their application is illustrated in Figure 9.

[5] https://www.w3.org/TR/prov-o/

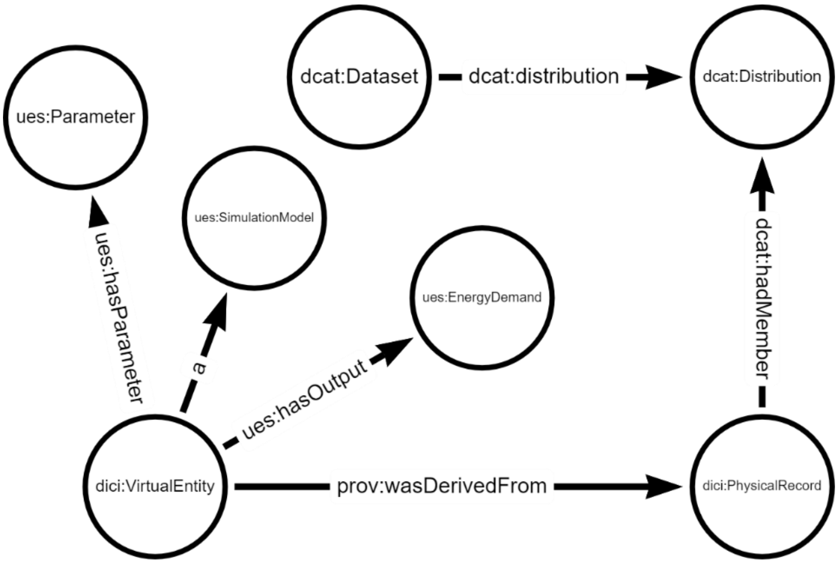


Figure 9: Demonstration of how the PROV-O ontology is used in the prototype to link multiple virtual entities to the record of a physical entity. The prefixes: ues: Urban Energy Systems ontology; dici: Digicites ontology; dcat: Datacatalog ontology; prov: PROV-O ontology.

4.4 RDF Triplestore

Example queries to extract data have been uploaded to the Digicities repository of the NEST Cloud[6]. Users can register for access to the data. (Note – the prototype is currently available of the NEST Cloud but in the future, this will be moved to the pilot platform).

Here are some example queries demonstrating how to explore and retrieve data from the platform. Figure 10 shows how to trace the original data source used to generate virtual building models.

```
PREFIX dici: <https://www.digicities.info/ontology/>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
select * where {
    ?building a dici:PhysicalBuildingRecord.
     ?virtual_building prov:wasDerivedFrom ?building;
                       prov:wasMemberOf ?data_set.
    ?data_set dcat:accessUrl ?original_source.
} limit 100
```

building	virtual_building	data_set	original_source
dici_bldg:CH_11141909	dici_bldg:CH_121773027_cesarp	dici_data:591fc6c0-e5cb-11ed-a05b-0242ac120003	dcat:10.5281/zenodo.7875922
dici_bldg:CH_11203716	dici_bldg:CH_121773028_cesarp	dici_data:591fc6c0-e5cb-11ed-a05b-0242ac120003	dcat:10.5281/zenodo.7875922
dici_bldg:CH_11141466	dici_bldg:CH_121773029_cesarp	dici_data:591fc6c0-e5cb-11ed-a05b-0242ac120003	dcat:10.5281/zenodo.7875922
dici_bldg:CH_11141888	dici_bldg:CH_121773031_cesarp	dici_data:591fc6c0-e5cb-11ed-a05b-0242ac120003	dcat:10.5281/zenodo.7875922

Figure 10: An example SPARQL query performed on the Digicities repository of the knowledge graph to return information regarding the provenance of virtual buildings generated from the data. Note the use of Uniform Resource Identifiers (URI) to indicate instances of entities stored on the platform

[6] https://graphdb.nestcloud.ch/login

Figure 11 shows how to extract specific demands of specific building types within a specific geospatial cell.

```
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX ues: <https://ja98.github.io/ues>
PREFIX dici_geo: <https://www.digicities.info/geometries/>
PREFIX dici: <https://www.digicities.info/ontology/>
select * where {
    ?virtual_building a dici:VirtualBuilding;
     geo:within dici_geo:h3_891f9d60c2fffff;
     ues:buildingUse ues:Office;
     ues:hasSimulatedCoolingDemand ?cooling_demand.
} limit 100
```

virtual_building	cooling_demand
dici_bldg:CH_138304080_cesarp	"29137.3905"^^<https://ja98.github.io/uesEnergyDemandPerYear>
dici_bldg:CH_138304106_cesarp	"16662.0324"^^<https://ja98.github.io/uesEnergyDemandPerYear>
dici_bldg:CH_138304115_cesarp	"13360.0853"^^<https://ja98.github.io/uesEnergyDemandPerYear>
dici_bldg:CH_138304123_cesarp	"22887.2395"^^<https://ja98.github.io/uesEnergyDemandPerYear>
dici_bldg:CH_138304132_cesarp	"25573.9184"^^<https://ja98.github.io/uesEnergyDemandPerYear>

Figure 11: An example SPARQL query performed on the Digicities repository of the NEST knowledge graph to return the cooling demand of virtual buildings, assigned as a ues:Office, within a specified H3 geospatial cell.

The querying may seem arbrirtray, however, if an application has prior knowledge of the ontologies and the terminological relationships, it will enable autonomous processing of the data that could improve the retrieval, processing and comprehension of the data. For example, the proposed structure could be useful in the development of natural language retrieval methods. A user would be shielded from the complexity of developing the complex queries but could ask to have all of the cooling demands for a region. The application would know handle this request based on the structure and linking of the data and provide the user a meaningful response with any underlying assumptions. This is currently not possible with data for urban demand forecasting so the approach in Digicities would have considerable benefit of the discovery and application of data for demand forecasting. The results can be queried according to the H3 geospatial ID. The kepler.gl web app[7] will automatically plot data that is organised accordign the the H3 ID – drag and drop interface. Files with a column with a H3 ID are automatically plotted and aggregated according to the value of interest. This process is shown in Figure 12.

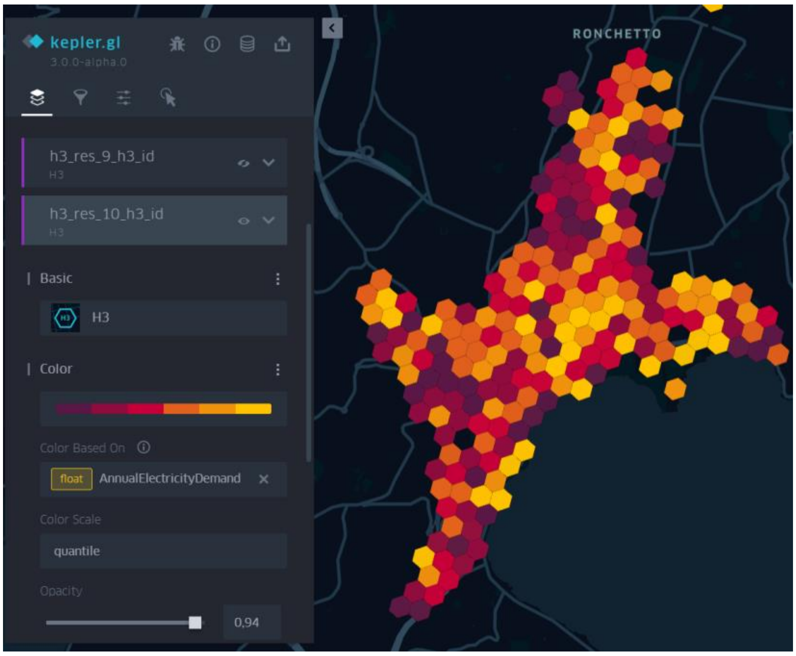


Figure 12: Visualisation of simulated annual electricity demand using H3 geospatial cells in kepler.gl

[7] https://kepler.gl/demo

4.5 Machine learning

IDSIA researchers have developed methods for energy load and peak forecasting placing them high in the BigDeal Challenge 2022. These methods are improved and detailed in a paper submitted to the 12th DACH+ conference on energy informatics (2023). They include a novel procedure for feature engineering and feature selection, based on cluster permutation of temperatures and calendar variables. Gradient boosting of trees capabilities are exploited and enhanced with trend modelling and probabilistic forecasts (see Figure 11). While most approaches focus on improving the accuracy of point prediction, models correctly quantifying the uncertainty of the predictions are necessary to support reliable decision making. Finally, an approach to forecast combination known as temporal hierarchies, is used to further improve the accuracy.

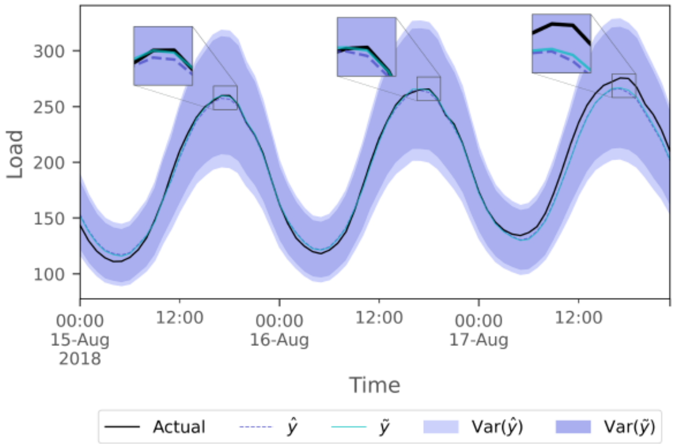


Figure 11: Comparison of probabilistic forecasts, before and after the application of the temporal hierarchy. The temporal hierarchy slightly improves the point forecasts. It also shortens the prediction intervals without compromising their reliability. Data are taken from the BigDeal Challenge 2022

Preliminary short term predictive models have been trained on data derived from the s the Lugaggia Innovation Community (LIC) project, a pilot promoting an experimental self-consumption energy community. Data includes smart-meters measures with a 15 minutes resolution, from 20 neighbouring houses with 5 photovoltaic plants and the distribution sub-station to which they are connected and 7 days ahead weather forecasts from Meteoblue provider. We used these data to train 15-minutes and 24 hours ahead forecasting models.

Concerning long term prediction based on building characteristics, we have started the analysis from data generated by EMPA simulator about annual heating, cooling, energy and hot water demand of Lugano buildings. We have built predictive models using different algorithms showing very high correlation of the prediction with the simulated values (see figure 12). Analysis of the predictor importance shows that the main contribution to such predictions is due to the building footprint area, as might have been expected. The high predictive performance observed are probably due also to the use of simulated data, which are probably generated by a simpler model than real ones. We plan to test the approach to real data in the next months.

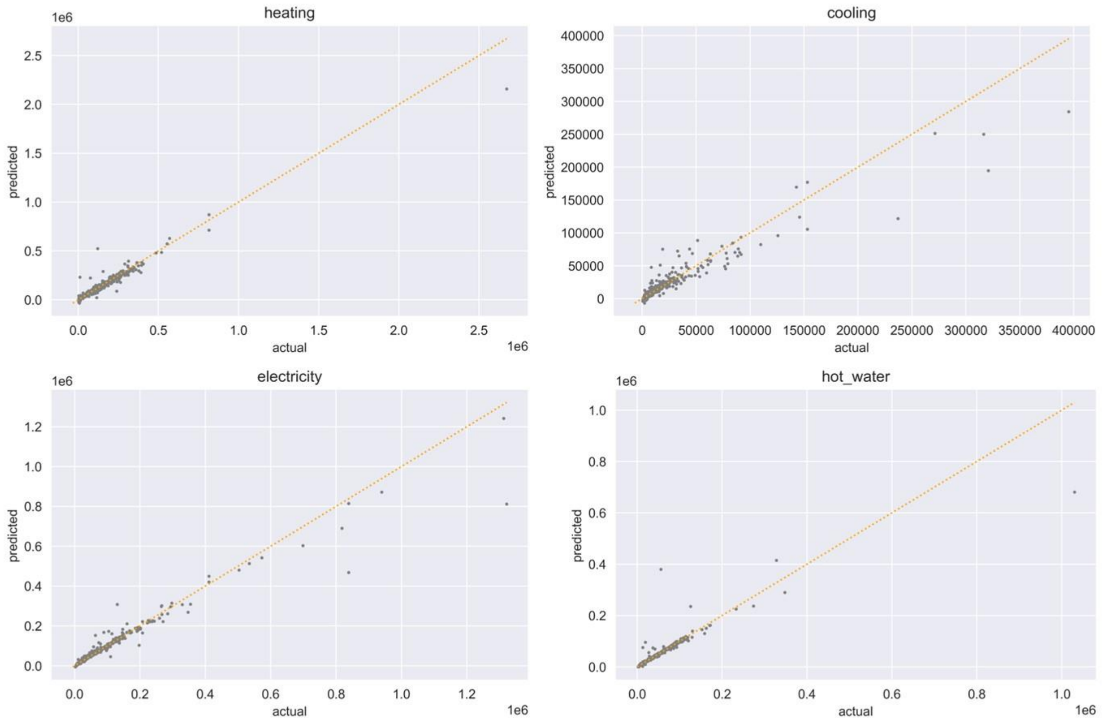


Figure 12: scatterplot of the demand forecasts versus the simulated values. The corresponding coefficients of determination are heating: R2 = 0.94, cooling: R2 = 0.88, electricity: R2 = 0.94, hot_water: R2 = 0.84

5 Next steps

An overview of the project phases and related tasks is shown in Figure 13.

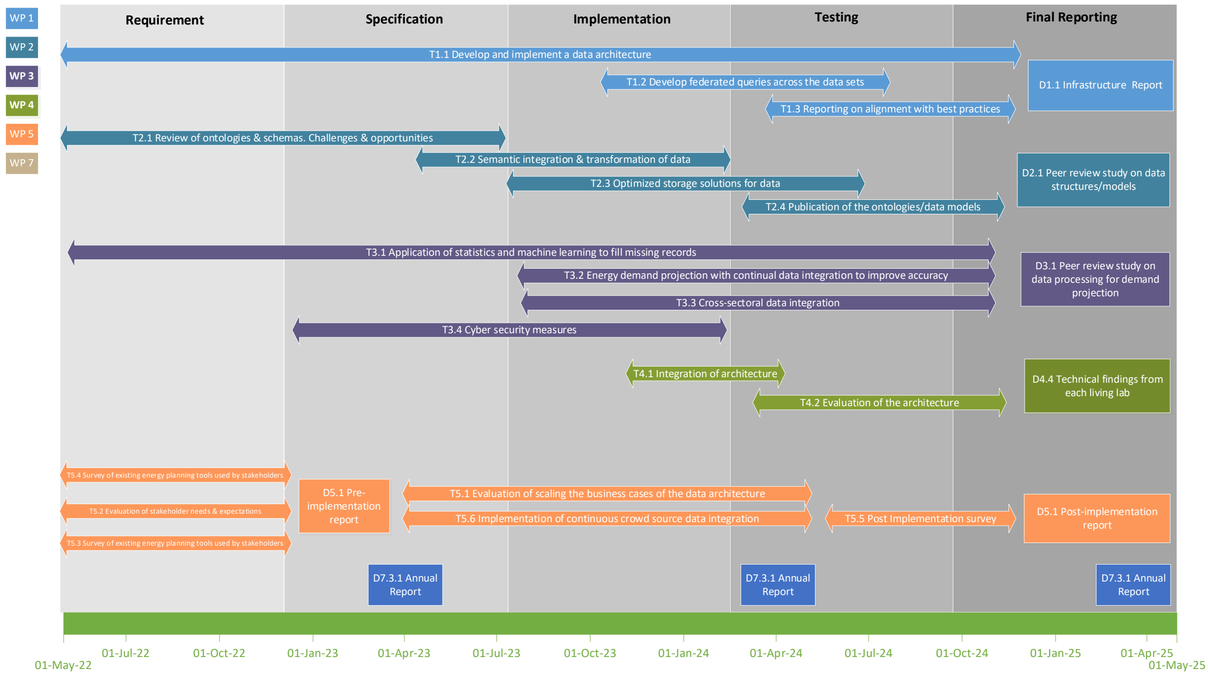


Figure 13: Project phases, tasks, and milestones

The project is entering the implementation phase with our Microsoft partners. We have devised a statement of work that will be financially supported by Microsoft. The formation of data contracts to satisfy the requirements of GAIA-X will be explored with TrustRelay. We have already had several calls and workshops but once we begin our implementation, we will align complimentary activities.

6 National and international cooperation

The Digicities project kicked-off on the 24th May 2022 in Lugano with 40 attendees from 20 organisations in three different countries (Switzerland, Austria, Spain). This included representatives from the need-owners and the living labs that will provide data to the project. The kick-off enabled us to build a knowledge community around our project that will be informed about major developments. A summary of those involved are shown on our webpage https://digicities.info/. An overview of the different project stakeholders is shown in Figure 14.

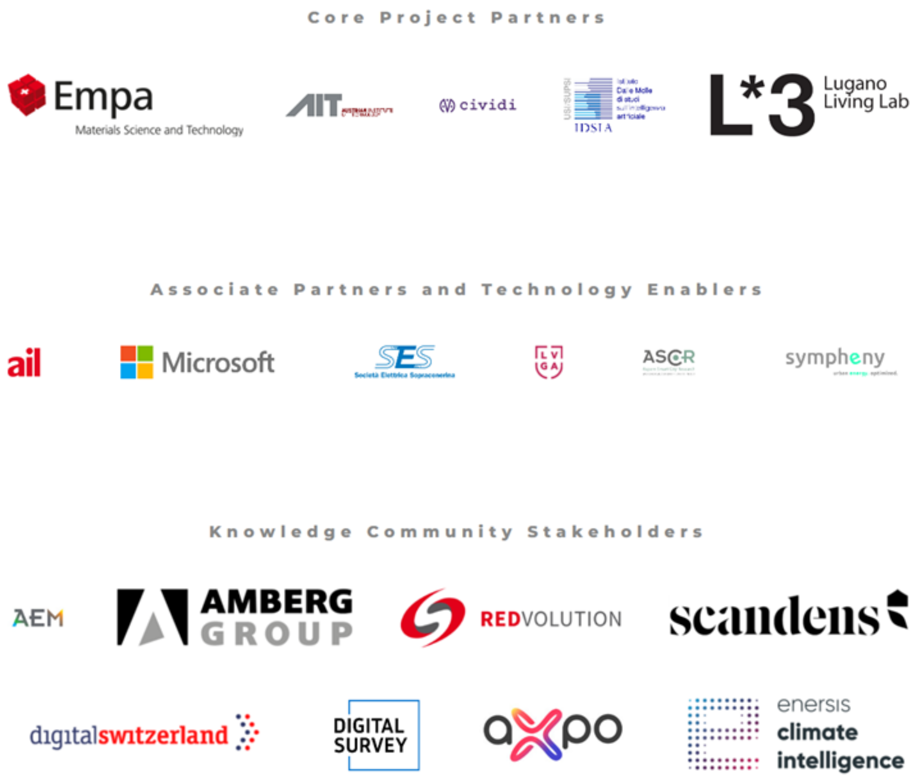


Figure 14: Digicities stakeholder network

The project is carried out in collaboration with partners in Austria. Their project is led by AIT who coordinate a use case focused on parameterised energy simulation models using the data exchange platform. They work with several regional partners and living labs. AIT recently coordinated and submitted a review that concluded in a set of requirements of the platform for their use case. They hope to present their paper at the DACH+ conference on Energy Informatics. A summary of the Austrian use case is shown in Figure 15.

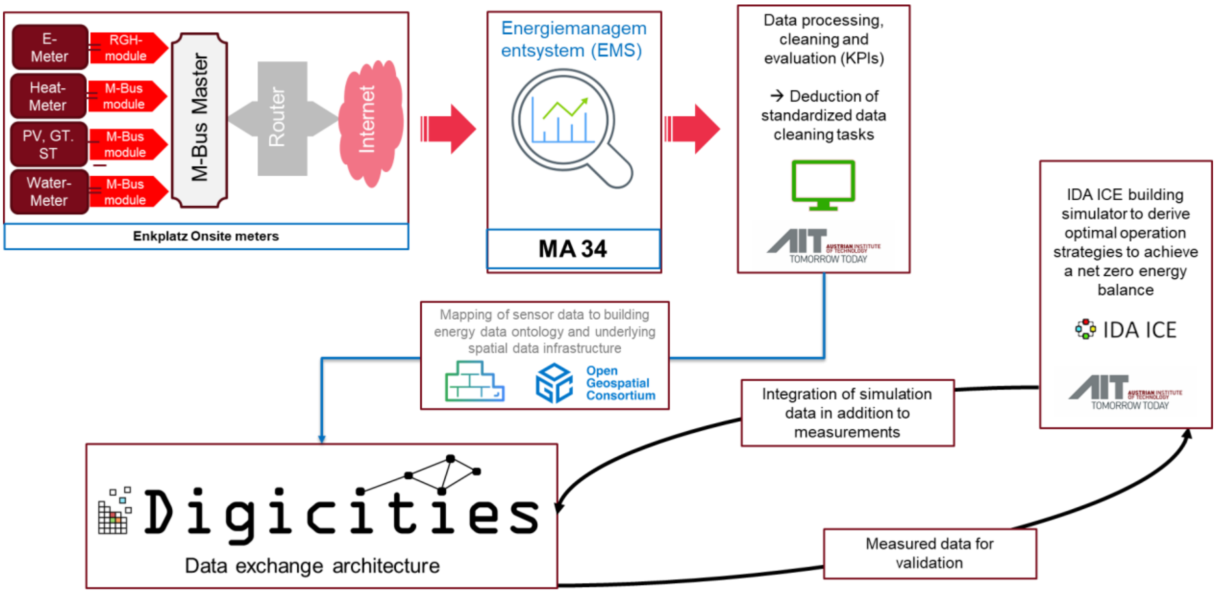


Figure 15: Austrian use case that will interface with the Digicities platform

7 Publications

Currently there are no publications; however, the following referenced manuscript will be published in the proceedings of the CISBAT 2023 conference:

Allan, J., Mangili, F., Derboni, M., Gisler, L., Hainoun, A., Rizzoli, A., Ventriglia, L., & Sulzer, M. (2023). A semantic data framework to support data-driven demand forecasting. In Proceedings of the CISBAT Conference (pp. XXX-XXX). Lausanne.

8 References

Agugiaro, G., Benner, J., Cipriano, P., Nouvel, R., 2018. The Energy Application Domain Extension for CityGML: enhancing interoperability for urban energy simulations. Open Geospatial Data Softw. Stand. 3, 1–30.

Ahmed, A., Khalid, M., 2019. A review on the selected applications of forecasting models in renewable power systems. Renew. Sustain. Energy Rev. 100, 9–21.

Allan, J., Fierz, L., Bollinger, A., Orehounig, K., 2021. Linked Data Ontology for Urban Scale Building Energy Simulation, in: ESIM 2020 Conference Proceedings.

Azure, M., 2022. Digital Twins Definition Language [WWW Document]. URL https://github.com/Azure/opendigitaltwins-dtdl/blob/b3f36cf3fb6062a9da68ed4b085a17c7a1144763/DTDL/v2/dtdlv2.md (accessed 9.28.22).

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Berges, M., Culler, D., Gupta, R., Kjærgaard, M.B., Srivastava, M., Whitehouse, K., 2016. Brick: Towards a Unified Metadata Schema For Buildings, in:

Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, BuildSys '16. Association for Computing Machinery, New York, NY, USA, pp. 41–50. https://doi.org/10.1145/2993422.2993577
Bastos Porsani, G., Del Valle de Lersundi, K., Sánchez-Ostiz Gutiérrez, A., Fernández Bandera, C., 2021. Interoperability between building information modelling (BIM) and building energy model (BEM). Appl. Sci. 11, 2167.
Battle, R., Kolas, D., 2011. Geosparql: enabling a geospatial semantic web. Semantic Web J. 3, 355–370.
BFS, 2020. Digital Switzerland Strategy [WWW Document]. URL https://www.bakom.admin.ch/dam/bakom/en/dokumente/informationsgesellschaft/strategie/strategie_digitale_schweiz.pdf.download.pdf/Strategie-DS-2020-EN.pdf
Biljecki, F., Kumar, K., Nagel, C., 2018. CityGML Application Domain Extension (ADE): overview of developments. Open Geospatial Data Softw. Stand. 3, 13. https://doi.org/10.1186/s40965-018-0055-6
Braud, A., Fromentoux, G., Radier, B., Le Grand, O., 2021. The road to European digital sovereignty with Gaia-X and IDSA. IEEE Netw. 35, 4–5.
Chadzynski, A., Krdzavac, N., Farazi, F., Lim, M.Q., Li, S., Grisiute, A., Herthogs, P., von Richthofen, A., Cairns, S., Kraft, M., 2021. Semantic 3D City Database—An enabler for a dynamic geospatial knowledge graph. Energy AI 6, 100106.
Cirillo, F., Solmaz, G., Berz, E.L., Bauer, M., Cheng, B., Kovacs, E., 2019. A standard-based open source IoT platform: FIWARE. IEEE Internet Things Mag. 2, 12–18.
Coakley, D., Raftery, P., Keane, M., 2014. A review of methods to match building energy simulation models to measured data. Renew. Sustain. Energy Rev. 37, 123–141.
Daniele, L., den Hartog, F., Roes, J., 2015. Created in close interaction with the industry: the smart appliances reference (SAREF) ontology. Presented at the Formal Ontologies Meet Industry: 7th International Workshop, FOMI 2015, Berlin, Germany, August 5, 2015, Proceedings 7, Springer, pp. 100–112.
Di Fonzo, T., Girolimetto, D., 2023. Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives. Int. J. Forecast. 39, 39–57.
Duffie, J.A., Beckman, W.A., Blair, N., 2020. Solar engineering of thermal processes, photovoltaics and wind. John Wiley & Sons.
Elagiry, M., Charbel, N., Bourreau, P., Di Angelis, E., Costa, A., 2020. IFC to building energy performance simulation: a systematic review of the main adopted tools and approaches. Proc. BauSIM.
Ferrero Bermejo, J., Gómez Fernández, J.F., Olivencia Polo, F., Crespo Márquez, A., 2019. A review of the use of artificial neural network models for energy and reliability prediction. A study of the solar PV, hydraulic and wind energy sources. Appl. Sci. 9, 1844.
Gantner, F., 2011. A spatiotemporal ontology for the administrative units of Switzerland.
Gantner, F., Waldvogel, B., Meile, R., Laube, P., 2013. The B asic F ormal O ntology as a Reference Framework for Modeling the Evolution of Administrative Units. Trans. GIS 17, 206–226.
Ghasempour, A., 2019. Internet of things in smart grid: Architecture, applications, services, key technologies, and challenges. Inventions 4, 22.
Gottschalk, M., Uslar, M., Delfs, C., Gottschalk, M., Uslar, M., Delfs, C., 2017. The smart grid architecture model–SGAM. Use Case Smart Grid Archit. Model Approach IEC 62559-2 Use Case Template SGAM Appl. Var. Domains 41–61.
Grütter, R., Purves, R.S., Wotruba, L., 2017. Evaluating topological queries in linked data using DBpedia and GeoNames in Switzerland and Scotland. Trans. GIS 21, 114–133.
Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M., 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Gener. Comput. Syst. 29, 1645–1660.
Guo, D., Onstein, E., 2020. State-of-the-art geospatial information processing in NoSQL databases. ISPRS Int. J. Geo-Inf. 9, 331.

H. Kolbe, T., Kutzner, T., Stephen Smyth, C., Nagel, C., Roensdorf, C., Heazel, C., 2021. OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard. Open Geospatial Consortium.

Habbak, H., Mahmoud, M., Metwally, K., Fouda, M.M., Ibrahem, M.I., 2023. Load Forecasting Techniques and Their Applications in Smart Grids. Energies 16, 1480.

Hammar, K., Wallin, E.O., Karlberg, P., Hälleberg, D., 2019. The RealEstateCore Ontology, in: Ghidini, C., Hartig, O., Maleshkova, M., Svátek, V., Cruz, I., Hogan, A., Song, J., Lefrançois, M., Gandon, F. (Eds.), The Semantic Web – ISWC 2019, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 130–145. https://doi.org/10.1007/978-3-030-30796-7_9

Hippolyte, J.-L., Rezgui, Y., Li, H., Jayan, B., Howell, S., 2018. Ontology-driven development of web services to support district energy applications. Autom. Constr. 86, 210–225.

Holles, S., Rütten, F., Türkucar, T., Oliver Thier, C., 2021. Datahub Schweiz. Kern zukünftigerDateninfra-strukturdigitaliserterStrom-und Gasmärkte. Bundesamt für Energie, Bern.

Hollyman, R., Petropoulos, F., Tipping, M.E., 2021. Understanding forecast reconciliation. Eur. J. Oper. Res. 294, 149–160.

Hong, T., Chen, Y., Luo, X., Luo, N., Lee, S.H., 2020. Ten questions on urban building energy modeling. Build. Environ. 168, 106508.

Hong, T., Fan, S., 2016. Probabilistic electric load forecasting: A tutorial review. Int. J. Forecast. 32, 914–938.

Kim, H.J., Jeong, C.M., Sohn, J.-M., Joo, J.-Y., Donde, V., Ko, Y., Yoon, Y.T., 2020. A comprehensive review of practical issues for interoperability using the common information model in smart grids. Energies 13, 1435.

Kolbe, T.H., Nagel, C., Herreruela, J., 2013. 3d city database for citygml. Add. 3D City Database Doc. Version 2.

Kutzner, T., Chaturvedi, K., Kolbe, T.H., 2020. CityGML 3.0: New Functions Open Up New Applications. PFG – J. Photogramm. Remote Sens. Geoinformation Sci. 88, 43–61. https://doi.org/10.1007/s41064-020-00095-z

Lambert, Eric, Boultadakis, George, Kukk, Kalle, Kotsalos, Konstantinos, Bilidis, Nikos, 2021. European energy data exchange reference architecture: Data Management Working Group.

Li, L., Ota, K., Dong, M., 2017. When weather matters: IoT-based electrical load forecasting for smart grid. IEEE Commun. Mag. 55, 46–51.

Liu, P., Lu, C., 2021. Strategic analysis and development plan design on digital transformation in the energy industry: A global perspective. Int. J. Energy Res. 45, 19657–19670. https://doi.org/10.1002/er.7032

Lu, Y., Tian, Z., Zhou, R., Liu, W., 2021. A general transfer learning-based framework for thermal load prediction in regional energy system. Energy 217, 119322.

Meira, E., Lila, M.F., Oliveira, F.L.C., 2023. A novel reconciliation approach for hierarchical electricity consumption forecasting based on resistant regression. Energy 269, 126794.

Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S., Varkonyi-Koczy, A.R., 2019. State of the art of machine learning models in energy systems, a systematic review. Energies 12, 1301.

Orehounig, K., Fierz, L., Allan, J., Eggimann, S., Vulic, N., Bojarski, A., 2022. CESAR-P: A dynamic urban building energy simulationtool. J. Open Source Softw. 7, 4261. https://doi.org/10.21105/joss.04261

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, 2024. https://doi.org/10.3390/en14072024

Qiu, R., Duan, H., Yu, X., Zheng, L., Cheng, Q., 2015. An ontology matching method for GML application ontology to GeoSPARQL ontology. Presented at the 2015 23rd International Conference on Geoinformatics, IEEE, pp. 1–5.

Salani, M., Derboni, M., Rivola, D., Medici, V., Nespoli, L., Rosato, F., Rizzoli, A.E., 2020. Non intrusive load monitoring for demand side management. Energy Inform. 3, 25. https://doi.org/10.1186/s42162-020-00128-2
Scheller, F., Bruckner, T., 2019. Energy system optimization at the municipal level: An analysis of modeling approaches and challenges. Renew. Sustain. Energy Rev. 105, 444–461. https://doi.org/10.1016/j.rser.2019.02.005
Schweizerische Bundesrat, 2008. Stromversorgungsverordnung, 734.71.
SFOE, S.F.O. of E., 2022. Energy Strategy 2050 [WWW Document]. URL https://www.bfe.admin.ch/bfe/en/home/politik/energiestrategie-2050.html (accessed 12.23.22).
Spiliotis, E., Petropoulos, F., Kourentzes, N., Assimakopoulos, V., 2020. Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. Appl. Energy 261, 114339.
Thirumuruganathan, S., Tang, N., Ouzzani, M., Doan, A., 2020. Data Curation with Deep Learning., in: EDBT. pp. 277–286.
University of Geneva, 2023. OWL version of the CityGML [WWW Document]. OWL Version CityGML. URL http://cui.unige.ch/isi/onto//citygml2.0.owl (accessed 3.27.23).
Wang, D., Landolt, J., Mavromatidis, G., Orehounig, K., Carmeliet, J., 2018. CESAR: A bottom-up building stock modelling tool for Switzerland to address sustainable energy transformation strategies. Energy Build. 169, 9–26.
Wetter, M., Zuo, W., Nouidui, T.S., Pang, X., 2014. Modelica buildings library. J. Build. Perform. Simul. 7, 253–270.
Zhu, J., Dong, H., Zheng, W., Li, S., Huang, Y., Xi, L., 2022. Review and prospect of data-driven techniques for load forecasting in integrated energy systems. Appl. Energy 321, 119269.