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

Final report dated 31st October 2025

DIGICITIES

Urban Digital Layers to Support the Energy Transition of Cities

Date: 31st October 2025

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/

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

Das Projekt DIGICITIES hat eine semantische Datenplattform entwickelt, um grundlegende Herausforderungen im urbanen Energiedatenmanagement zu bewältigen: fragmentierte Datensätze, inkonsistente Terminologie, arbeitsintensive Datenaufbereitung und mangelnde Nachvollziehbarkeit von Annahmen in Energieplanungsstudien. Konventionelle Ansätze erfordern von Energieplanern und Dienstleistern die manuelle Integration heterogener Datenquellen, die Zuordnung proprietärer Formate und wiederholte Datentransformationen für jedes Analysewerkzeug. Dieser Prozess ist fehleranfällig, zeitaufwändig und schwer reproduzierbar oder überprüfbar.

Innovation und Funktionsprinzip der Plattform. Die Plattform führt drei zentrale Innovationen ein. Erstens eine **flexible, dienstleistungsorientierte Ontologie**, die den traditionellen Ansatz zur Standardadoption umkehrt. Anstatt von Entwicklern zu verlangen, grosse, komplexe Domänenontologien wie CIM oder SAREF vollständig zu übernehmen, ermöglicht DIGICITIES Dienstleistern, leichtgewichtige Erweiterungen zu definieren, die nur die für ihre Anwendungen erforderlichen Komponenten und Attribute spezifizieren, während eine systematische Zuordnung zu etablierten Standards die Interoperabilität bei Bedarf gewährleistet. Zweitens kombinieren **semantische Datenprodukte** maschinenlesbare RDF-Beschreibungen mit zugehörigen Ressourcendateien (CSV-Zeitreihen, GeoJSON-Geometrien, EPW-Wetterdaten) zu eigenständigen, auffindbaren Paketen, die Mehrdeutigkeiten hinsichtlich Dateninhalt, Einheiten und Herkunft beseitigen. Drittens ermöglicht das **Szenariomanagement mit Rückverfolgbarkeit der Annahmen** Planern, Szenariovarianten zu erstellen, zu modifizieren und zu vergleichen, wobei eine vollständige Dokumentation der geänderten Annahmen erhalten bleibt. Zusammen reduzieren diese Funktionen den manuellen Datenintegrationsaufwand, der derzeit die Arbeitsabläufe in der Energieplanung dominiert, und schaffen einen standardisierten Mechanismus für den Datenaustausch entlang der digitalen Wertschöpfungskette.

Anwendungsfälle und Testmethodik. Die Plattform wurde anhand von zwei unterschiedlichen Anwendungsfällen validiert. **UC1 (Operative Entscheidungsfindung)** setzte modellprädiktive Regelung (MPC) ein, um die Einsatzplanung erneuerbarer Energieressourcen für zwei Wasserkraftwerke (AEM und SES) sowie eine Eigenverbrauchsgemeinschaft aus 10 Haushalten mit Batteriespeicher zu optimieren. LightGBM-basierte Prognosemodelle wurden mit historischen Nachfrage-, Zufluss- und Wetterdaten trainiert, um 48-Stunden-Prognosen zu erstellen. Ein Scheduler optimierte anschliessend die Produktionspläne zur Minimierung der monatlichen Lastspitzen. Die Leistung wurde bewertet, indem die simulierten Spitzenreduktionen des Controllers mit den tatsächlichen Entscheidungen der Betreiber über mehrere monatliche Testperioden verglichen wurden. Für die Batteriedimensionierungsstudie wurden systematische Kombinationen von PV-Kapazität (51–205 kWp) und Batteriekapazität (0–200 kWh) über drei saisonale Monate simuliert. **UC2 (Energieplanung)** konzentrierte sich auf die kommunale Energieberichterstattung und szenariobasierte Planung. Bestehende Energieplanungsdaten und Modellkonfigurationen wurden in semantische Datenprodukte umgewandelt und über den Szenario-Builder der Plattform integriert. Eine regionale Energieplanungs-API (Sympheny) wurde angebunden, um die direkte Einreichung von auf der Plattform erstellten Szenarien zu ermöglichen.

Quantitative Ergebnisse. In UC1 übertraf der LightGBM-basierte Scheduler sowohl Basismodelle als auch erfahrene menschliche Betreiber bei der Spitzenlastreduzierung durchgehend. Für AEM wurde eine durchschnittliche monatliche Einsparung von ca. 16 kCHF über vier Testperioden geschätzt. Für SES lagen die Einsparungen zwischen 96 und 220 kCHF/Monat je nach Planungsfrequenz (täglich vs. alle 15 Minuten), entsprechend Spitzenreduktionen von 4–8 MW. Die Batteriedimensionierungssimulationen zeigten, dass eine Kombination aus grosser PV-Anlage (205 kWp) und grossem Speicher (100 kWh) eine vollständige tägliche Batterieauslastung mit einem Autarkiegrad von 48–78% je nach Saison erreichte, während ein übergrosser Speicher (200 kWh) bis zu 91% Autarkie im Sommer erzielte,

jedoch mit abnehmenden Erträgen. Die Prognosegenauigkeit erwies sich als entscheidend: Das LGBM-Modell vermied fehlerhafte Batterieladungen aus dem Netz im Winter, die bei einfacheren Modellen beobachtet wurden. Für UC2 erzeugte die Plattform erfolgreich neue Datenprodukte für E-Ladestationen- und Solarfassadenpotenzial und ermöglichte die direkte API-Einreichung an den regionalen Planungsdienst Sympheny, auch wenn quantifizierte Zeiteinsparungen im Projektzeitraum nicht formal evaluiert werden konnten.

Schlussfolgerungen zu Nutzen und Wirkung der Plattform. Das Projekt zeigt, dass eine semantische Dateninfrastruktur messbare betriebliche Vorteile im Energieressourcenmanagement liefern kann und eine technisch tragfähige Grundlage für den standardisierten Datenaustausch in der städtischen Energieplanung bietet. Die Ergebnisse der Wasserkraftsteuerung bestätigen, dass eine automatisierte, prognosegestützte Steuerung durch strukturierte Datenpipelines die manuelle Expertenplanung übertreffen kann, mit finanziell signifikanten Spitzenreduktionen. Digitale Dienstleister identifizierten klare Vorteile, darunter reduzierter Datenintegrationsaufwand, verbesserte API-Datenqualität und skalierbare Szenarienbearbeitung. Die Stakeholder-Bewertung ergab jedoch auch, dass Energieplaner und Gemeinden Migrationskosten und Arbeitsablaufunterbrechungen als primäre Hindernisse betrachten, wobei konkrete Nachweise durch zusätzliche Fallstudien und quantifizierte Zeiteinsparungen als wesentliche Voraussetzungen für Adoptionsentscheidungen identifiziert wurden. Die langfristige Wirkung der Plattform hängt daher von der Erweiterung des Katalogs gebrauchsfertiger semantischer Datenprodukte, der Demonstration des Nutzens durch zunehmend komplexe Praxisanwendungen und der Abstimmung mit entstehenden Datenraum-Initiativen wie GAIA-X und der Swiss Data Alliance ab, um die Governance- und Marktbedingungen zu schaffen, die eine Teilnahme am standardisierten Energiedatenaustausch fördern.

Résumé

Le projet DIGICITIES a développé une plateforme de données sémantiques pour répondre aux défis fondamentaux de la gestion des données énergétiques urbaines : jeux de données fragmentés, terminologie incohérente, préparation des données laborieuse et traçabilité limitée des hypothèses dans les études de planification énergétique. Les approches conventionnelles exigent des planificateurs et des prestataires l'intégration manuelle de sources de données hétérogènes, la correspondance de formats propriétaires et la transformation répétée des données pour chaque outil analytique. Ce processus est sujet aux erreurs, chronophage et difficilement reproductible ou vérifiable.

Innovation et principe de fonctionnement de la plateforme. La plateforme introduit trois innovations principales. Premièrement, une **ontologie flexible orientée services** qui inverse l'approche traditionnelle d'adoption des standards. Plutôt que d'exiger des développeurs l'adoption complète d'ontologies de domaine complexes telles que CIM ou SAREF, DIGICITIES permet aux prestataires de définir des extensions légères spécifiant uniquement les composants et attributs requis par leurs applications, tout en préservant l'interopérabilité grâce à un système de correspondance systématique avec les standards établis. Deuxièmement, les **produits de données sémantiques** combinent des descriptions RDF lisibles par machine avec des fichiers de ressources associés (séries temporelles CSV, géométries GeoJSON, données météorologiques EPW) en paquets autonomes et découvrables qui éliminent toute ambiguïté sur le contenu, les unités et la provenance des données. Troisièmement, la **gestion de scénarios avec traçabilité des hypothèses** permet aux planificateurs de construire, modifier et comparer des variantes de scénarios tout en conservant un historique complet des hypothèses modifiées. Ensemble, ces fonctionnalités réduisent la charge d'intégration manuelle des données qui domine actuellement les flux de travail en planification énergétique et créent un mécanisme standardisé pour l'échange de données le long de la chaîne de valeur numérique.

Cas d'utilisation et méthodologie de test. La plateforme a été validée à travers deux cas d'utilisation distincts. **UC1 (Prise de décision opérationnelle)** a appliqué une commande prédictive par modèle (MPC) pour optimiser la planification des ressources énergétiques renouvelables pour deux centrales hydroélectriques (AEM et SES) et une communauté d'autoconsommation solaire de 10 ménages avec stockage par batterie. Des modèles de prévision basés sur LightGBM ont été entraînés sur des données historiques de demande, d'afflux d'eau et de météo pour produire des prévisions à 48 heures. Un planificateur a ensuite optimisé les plans de production pour minimiser les pointes de demande mensuelles. La performance a été évaluée en comparant les réductions de pointe simulées du contrôleur aux décisions réelles des opérateurs sur plusieurs périodes de test mensuelles. Pour l'étude de dimensionnement des batteries, des combinaisons systématiques de capacité PV (51–205 kWc) et de batterie (0–200 kWh) ont été simulées sur trois mois saisonniers. **UC2 (Planification énergétique)** s'est concentré sur le rapportage énergétique municipal et la planification par scénarios. Les données et configurations de modèles existantes ont été transformées en produits de données sémantiques et intégrées via le constructeur de scénarios de la plateforme. Une API de planification énergétique régionale (Sympheny) a été connectée pour permettre la soumission directe des scénarios construits sur la plateforme.

Résultats quantitatifs. Dans UC1, le planificateur basé sur LightGBM a systématiquement surpassé les modèles de référence et les opérateurs humains expérimentés en matière de réduction des pointes de demande. Pour AEM, une économie mensuelle moyenne d'environ 16 kCHF a été estimée sur quatre périodes de test. Pour SES, les économies se situaient entre 96 et 220 kCHF/mois selon la fréquence de planification (quotidienne vs. toutes les 15 minutes), correspondant à des réductions de pointe de 4–8 MW. Les simulations de dimensionnement de batterie ont montré qu'une combinaison grande installation PV (205 kWc) et grande batterie (100 kWh) atteignait une utilisation quotidienne

complète de la batterie avec un taux d'autosuffisance de 48–78% selon la saison, tandis qu'une très grande batterie (200 kWh) atteignait jusqu'à 91% d'autosuffisance en été mais avec des rendements décroissants. La précision des prévisions s'est avérée cruciale : le modèle LGBM a évité les erreurs de charge de batterie depuis le réseau en hiver observées avec des modèles plus simples. Pour UC2, la plateforme a généré avec succès de nouveaux produits de données pour le potentiel de recharge de véhicules électriques et de façades solaires et a permis la soumission directe via API au service Sympheny, bien que les gains de temps quantifiés n'aient pu être formellement évalués dans le cadre temporel du projet.

Conclusions sur le mérite et l'impact de la plateforme. Le projet démontre qu'une infrastructure de données sémantiques peut fournir des avantages opérationnels mesurables dans la gestion des ressources énergétiques et constitue une base techniquement viable pour l'échange standardisé de données dans la planification énergétique urbaine. Les résultats de la planification hydroélectrique confirment qu'un contrôle automatisé piloté par des prévisions, rendu possible par des pipelines de données structurés, peut surpasser la planification manuelle d'experts, avec des réductions de pointe financièrement significatives. Les prestataires de services numériques ont identifié des avantages clairs, notamment la réduction de la charge d'intégration des données, l'amélioration de la qualité des données API et la gestion multi-scénarios à grande échelle. Cependant, l'évaluation des parties prenantes a également révélé que les planificateurs énergétiques et les municipalités considèrent les coûts de migration et la perturbation des flux de travail comme les principaux obstacles, des preuves concrètes à travers des études de cas supplémentaires et des gains de temps quantifiés étant identifiés comme des prérequis essentiels. L'impact à long terme de la plateforme dépend donc de l'extension du catalogue de produits de données sémantiques prêts à l'emploi, de la démonstration de la valeur à travers des applications réelles de complexité croissante, et de l'alignement avec les cadres émergents d'espaces de données tels que GAIA-X et la Swiss Data Alliance pour créer les conditions de gouvernance et de marché qui encouragent la participation à l'échange standardisé de données énergétiques.

Executive Summary

The DIGICITIES project developed a semantic data platform to address fundamental challenges in urban energy data management: fragmented datasets, inconsistent terminology, labour-intensive data preparation, and limited traceability of assumptions in energy planning studies. Conventional approaches require energy planners and service providers to manually integrate heterogeneous data sources, map proprietary formats, and repeatedly transform data for each analytical tool. This process is error-prone, time-consuming, and difficult to reproduce or audit.

Platform Innovation and Working Principle. The platform introduces three core innovations. First, a **flexible, service-driven ontology** that inverts the traditional approach to standards adoption. Rather than requiring developers to adopt large, complex domain ontologies such as CIM or SAREF in their entirety, DIGICITIES allows service providers to define lightweight extensions specifying only the components and attributes their applications require, while systematic mapping to established standards preserves interoperability when needed. Second, **semantic data products** combine machine-readable RDF descriptions with associated resource files (CSV time series, GeoJSON geometries, EPW weather data) into self-contained, discoverable packages that eliminate ambiguity about data content, units, and provenance. Third, **scenario management with assumption traceability** enables planners to construct, modify, and compare scenario variants while maintaining a full record of which assumptions were changed and why. Together, these features reduce the manual data integration burden that currently dominates energy planning workflows and create a standardised mechanism for data exchange across the digital value chain.

Use Cases and Testing Methodology. The platform was validated through two distinct use cases. **UC1 (Operational Decision Making)** applied model predictive control (MPC) to optimise renewable energy resource scheduling for two hydropower plants (AEM and SES) and a 10-household solar self-consumption community with battery storage. LightGBM-based forecasting models were trained on historical demand, water inflow, and weather data to produce 48-hour demand and inflow forecasts. A scheduler then optimised production plans to minimise monthly peak demand. Performance was evaluated by comparing the controller’s simulated peak reductions against actual operator decisions over multiple monthly test periods. For the battery sizing study, systematic combinations of PV capacity (51–205 kWp) and battery capacity (0–200 kWh) were simulated across three seasonal months to quantify self-consumption and self-sufficiency trade-offs. **UC2 (Energy Planning)** focused on municipal energy reporting and scenario-based planning. Existing energy planning data and model configurations from a utility and a municipal planner were transformed into semantic data products and integrated through the platform’s scenario builder. A regional energy planning API (Sympheny) was connected to enable direct submission of scenarios constructed on the platform.

Quantitative Results. In UC1, the LightGBM-based scheduler consistently outperformed both baseline models and experienced human operators in peak demand reduction. For AEM, an average monthly saving of approximately 16 kCHF was estimated over four test periods. For SES, savings ranged from 96 to 220 kCHF/month depending on scheduling frequency (daily vs. every 15 minutes), corresponding to peak reductions of 4–8 MW. The battery sizing simulations demonstrated that a Large PV (205 kWp) and Large Battery (100 kWh) combination achieved full daily battery utilisation with self-sufficiency reaching 48–78% depending on season, while a Very Large Battery (200 kWh) achieved up to 91% self-sufficiency in summer but with diminishing returns indicating over-sizing. Forecasting accuracy was critical: the LGBM model avoided erroneous winter battery charging from the grid observed with simpler models. For UC2, the platform successfully generated new data products for EV charging and solar façade potential and enabled direct API submission to the Sympheny regional planning service, though quantified time savings could not be formally evaluated within the project timeline.

Conclusions on Platform Merit and Impact. The project demonstrates that semantic data infrastructure can deliver measurable operational benefits in energy resource management and provides a technically viable foundation for standardised data exchange in urban energy planning. The HPP scheduling results confirm that automated, forecast-driven control enabled by structured data pipelines can outperform manual expert scheduling, with financially significant peak reductions. Digital service providers identified clear advantages including reduced data integration workloads, improved API data quality, and scalable multi-scenario handling. However, the stakeholder assessment also revealed that energy planners and municipalities view migration costs and workflow disruption as primary barriers, with concrete proof through additional case studies and quantified time savings identified as essential prerequisites before adoption decisions. The platform’s long-term impact therefore depends on expanding the catalog of ready-to-use semantic data products, demonstrating value through increasingly complex real-world applications, and aligning with emerging data space frameworks such as GAIA-X and the Swiss Data Alliance to create the governance and market conditions that incentivise participation in standardised energy data exchange.

Main findings

1 **Flexible Ontology Addresses Standard Adoption Barriers**: The DIGICITIES ontology allows service developers to create lightweight extensions to meet their needs while also giving the flexibility to map to established standards like CIM, when necessary. This service-driven approach reduces complexity while maintaining interoperability through mapping to established standards when necessary.

2 **Semantic Data Products Enable Standardized Exchange:** The semantic data product structure, combining semantic graphs with resource files (CSV, GeoJSON, EPW), provides a standardized publication mechanism that eliminates ambiguity and encourages interoperability of services. The option to ingest data products reduces data preparation overhead for consumers and ensures publishers create discoverable, interpretable datasets suitable for direct ingestion into scenario workflows.

3 **Multi-Stakeholder Value Chain Connection Demonstrated:** The platform successfully connects data publishers, service providers, and end users with tooling spanning ontology management, data product cataloging, scenario building, and API submission.

4 **Service Provider Benefits:** Digital service providers identified substantial advantages including reduced data integration workload, accelerated user data preparation, expanded market reach through simplified interfaces, increased confidence in incoming API data quality, and improved efficiency handling multiple scenarios.

5 **Barriers to Adoption for End Users:** Energy planners and municipal stakeholders viewed migration challenges and workflow transition costs as primary barriers to adoption, with time required for implementation currently outweighing perceived benefits. Concrete proof through additional case studies and quantified time savings are essential prerequisites before commitment decisions can be made.

6 **Improvements to Data Infrastructure Appeals More Than Modeling Efficiency**: Stakeholder interest focused primarily on standardized terminology, data ingestion improvements, traceability mechanisms, and workspace management capabilities rather than computational efficiency gains. This suggests positioning should emphasize data governance and quality assurance benefits when demonstrating the value through more complex projects.

7 **Successful Use Case Implementation:** The HPP scheduling application demonstrated feasibility of automated renewable resource optimization, achieving estimated monthly savings of 16 kCHF/month and between 96-220kCHF/month through peak reduction for the two use cases considered using peak demand reduction.

8 **Decentralized Architecture Aligns with Data Space Principles**: Future evolution toward federated storage where users connect their own cloud infrastructure would transition from centralized data custody to distributed sovereignty, with the platform operating as computational orchestration rather than data custodian. This aligns with IDSA and European Data Strategy frameworks emphasizing owner-controlled data exchange.

9 **Ecosystem Development Determines Long-Term Success:** Achieving transformative potential requires expanding semantic data product catalogs providing immediately usable datasets and diversifying the service offering that could bring value to the need owners of each use case.

10 **Concrete Applications Required to Overcome Perceived Barriers**: Converting existing resources (technology databases, demand profiles, building datasets) into standardized data products and applying the platform's tools in complex energy planning studies are critical next steps. Without demonstrated value through such applications, it will be challenging to gain widespread adoption in the industry.

Abbreviations

ADTP	Azure Digital Twin Platform
AEM	Azienda Elettrica di Massagno
AET	Azienda Elettrica Ticinese
API	Application Programming Interface
DTDL	Digital Twin Definition Language
ETL	Extract, Transform, Load
GWR	Gebäude und Wohnungsregister (Federal Register of Buildings and Dwellings)
HPP	Hydro Power Plant
KPI	Key Performance Indicator
LCA	Life-cycle analysis
MPC	Model Predictive Control
RDF	Resource Description Framework
RMSE	Root Mean Square Error
sMAPE	Symmetric Mean Absolute Percentage Error
UC	Use case
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
MVP	Minimum Viable Product
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

Table of Contents

1 Introduction

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, more 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 rapidly increasing the volume of data generated and available.

1.2 Energy Data Exchange 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 grid operators handle the growing complexity and requirements; however, the report mentioned a need to improve access and use of data significantly. 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 recommends standardising data and using 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).

1.3 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 usually 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 diverse 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, specific 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. Real-time integration of data from sensors and meters is becoming increasingly important. The rapidly growing and evolving Internet of Things (IoT) sector makes data collection from such devices possible. IoT refers to the interconnection of physical devices to the Internet, such as sensors and systems. It enables them to collect, exchange and analyse data to enable new capabilities for applications

---

[1] https://opendata.swiss

spanning numerous domains (Gubbi et al., 2013). In the energy sector, IoT can support the implementation of smart grids across all stages of the 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.

1.4 Data Availability Challenges

The challenge of data exchange in the energy sector is not a problem 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 on adopting common information models in the energy sector. The report refers to the Smart Grid Architecture Model (SGAM), which organises 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 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 increasingly use 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 is not limited to CIM but includes: 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 for data exchange 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 digitalising the energy infrastructure (Pritoni et al., 2021).

1.5 Data Quality Challenges

Data Cleaning includes detecting and repairing errors in a dataset to improve the data quality. This includes qualitative data cleaning, which uses 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.

1.6 Importance of Load Forecasting

Load forecasting predicts future energy demand using historical data and 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 on forecasts from a range of time horizons spanning seconds to hours, for demand response, days to months for energy trading, and years or decades for system planning and strategic policy-making (Hong and Fan, 2016). Demand forecasting can also be used as inputs to energy system optimisation models to support 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).

1.6.1 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 occurs 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 accurately represent real-world building operations (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 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). Including 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).

1.6.2 Load Forecasting with Machine-learning

Many machine-learning approaches have been proposed in recent years in response to the increased availability of data, particularly smart grid monitoring data. These methods have been shown to achieve a high accuracy. However, they have seldom been studied in a real or realistic framework. High accuracy or AUC (Area under the ROC Curve) on benchmark data does not necessarily result in effectiveness in real situations. Some requirements may be overlooked in the context of ML development. For instance, awareness of the uncertainty of the forecasts produced may be useful for 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 the 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 also be 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).

2 Project Goals

The aim of DIGICITIES was to develop, implement, and demonstrate a data exchange platform that connected different need-owners across the digital value chain. These were classified as data publishers, service providers and end-users. A set of use cases was defined by surveys with these groups of need-owners. The development addressed several of the recommendations for a data hub (Holles et al., 2021): simplifying data structures, use cases for data access and innovation, and extending the existing data architecture to consider new data sources. The aim was to demonstrate how cross-sector data can be mapped using standardised models to predict the energy demand at a specified output level. The platform was designed to supply parameters for both UBEM and ML studies. The critical functionality of the platform was to provide structured semantic information relevant to each use case. This project aimed to demonstrate how the datasets can be applied for energy demand forecasting using ML models or UBEM. This project aligned with the objective of increased renewables in the energy strategy (Swiss Federal Office of Energy, 2022), where models evaluate the impact of intermittent generation from renewables and plan according to the digital strategy for Switzerland (BFS, 2020).

2.1 Original 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

2.2 Planned Outputs

- Generation of additional features to train machine learning models
- Guidelines for 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

2.3 Research Questions

In addition to the above objectives, the project aimed to answer the following research questions:

2.3.1 **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 under-utilised data resources 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 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 urban data sources 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 integrating 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?

2.3.2 **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?

2.3.3 **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?

3 Background

3.1 Use Case Surveys

The purpose of profiling the inputs and outputs of our digital services was to understand the types of data they need and the value they can generate from it. This enabled a common understanding of requirements for preparing and exchanging data by the data publishers and service providers.

3.1.1 Survey of Usage

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. 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 Requirements

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

Organisations offer Digital services 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?*

3.2 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.

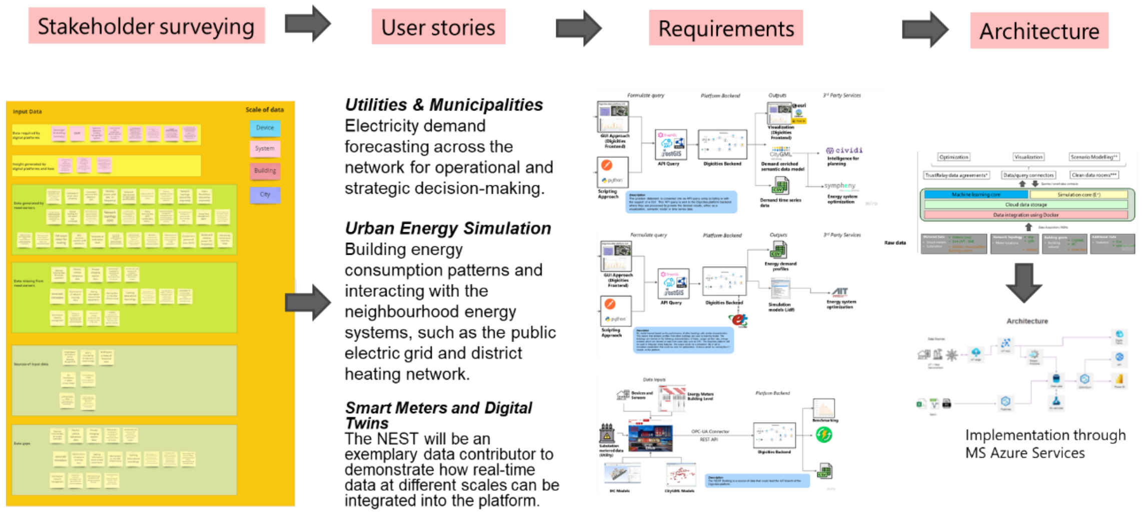


Figure 1: The process to design the conceptual architecture of the DIGICITIES platform. The user stories were generated from the results of a stakeholder survey were analysed.

Once the architecture was mapped out, it was sent to our implementation partner to identify the cloud services that could meet the requirements.

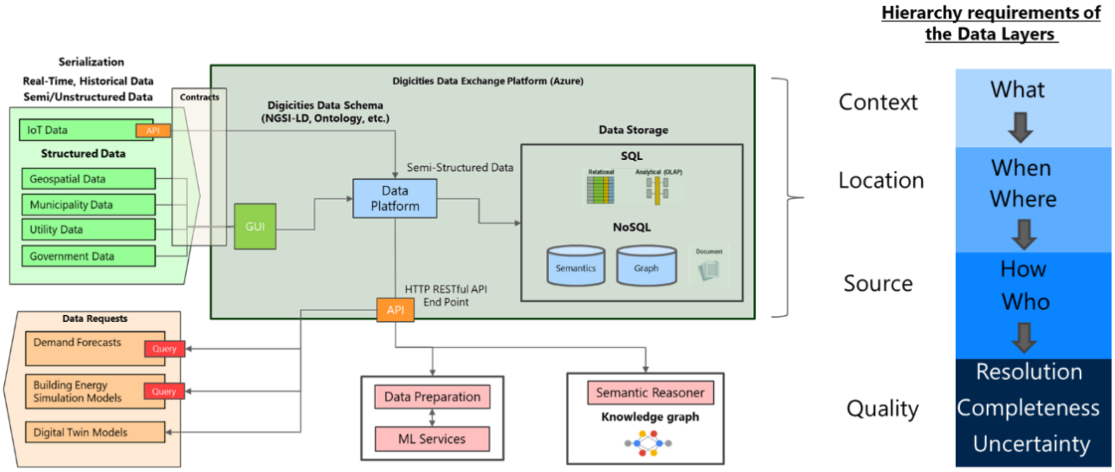


Figure 2: A more detailed look into the conceptual architecture covering specific data contributions and requests. The DIGICITIES data platform handles the data requests and contributions. The platform has a data storage ecosystem that enables data to be stored optimally according to the data format 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 Classification of Available Data

3.3.1 Data Interactions

The original conceptual platform had two forms of interaction: contributions and requests. The business case for closing the loop by providing services enabled by data enhanced by the DIGICITIES platform is evaluated in this project. An overview of the interactions is shown in Figure 3.

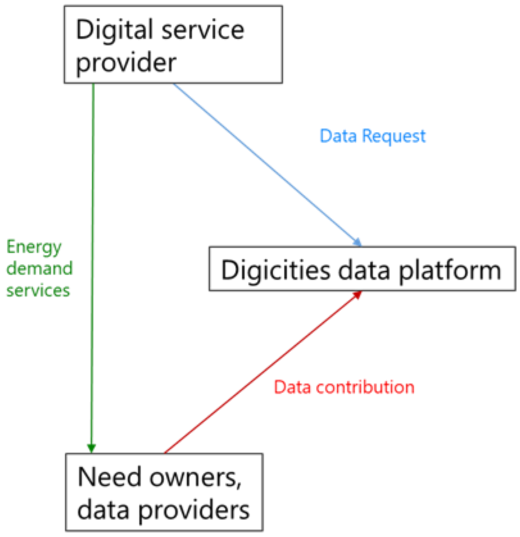


Figure 3: Data interactions foreseen in the DIGICITIES project

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 is shown in Figure 4.

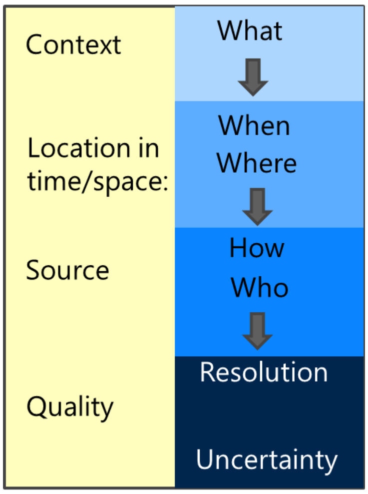
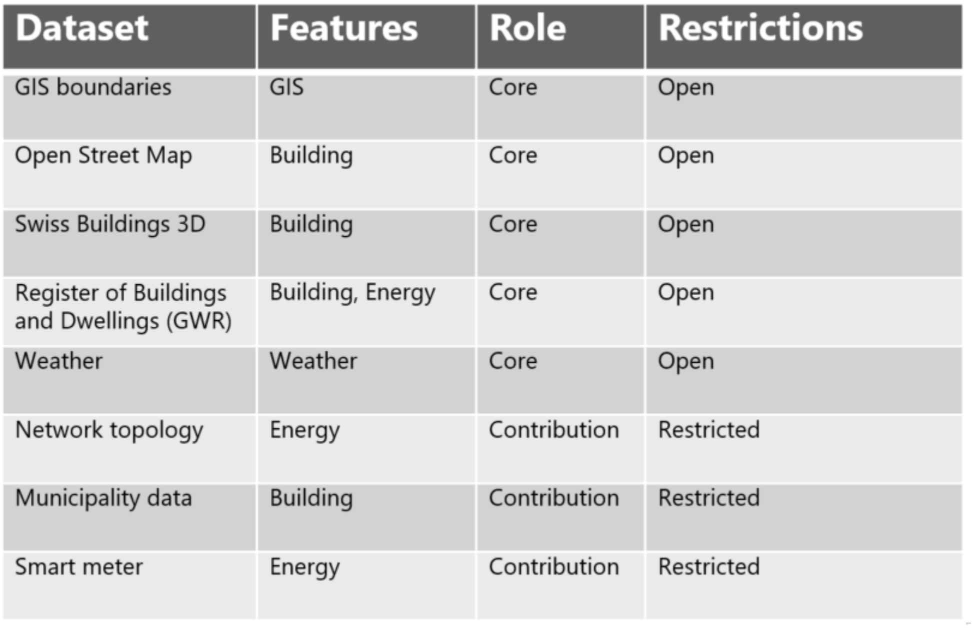


Figure 4: Semantic requirements of the data layers in the platform

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 a software system that works 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 that can 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. In contrast, graph and key-value databases are limited in data structures and 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 with contributions from many organisations in each country. The GeoNames Ontology is designed to enable the use of geospatial information on the semantic web[2].

---

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

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. 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

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.

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 integrate 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 a leading reference standard for exchanging BIM models. However, it is a very large and complex model, and there can be inconsistency in application (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 limited classes 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 unnecessary for the IFC, there are interoperability issues with the IFC BIM format and the gbXML BEM model (Bastos Porsani et al., 2021).

3.5.5 Energy and Timeseries

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 developing the ASHRAE 223p dictionary. While the development does not seem to be finished, it is aimed to be adopted as an 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 providers and 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 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 coordinating 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).

Figure 5 shows the data models reviewed and their classification.

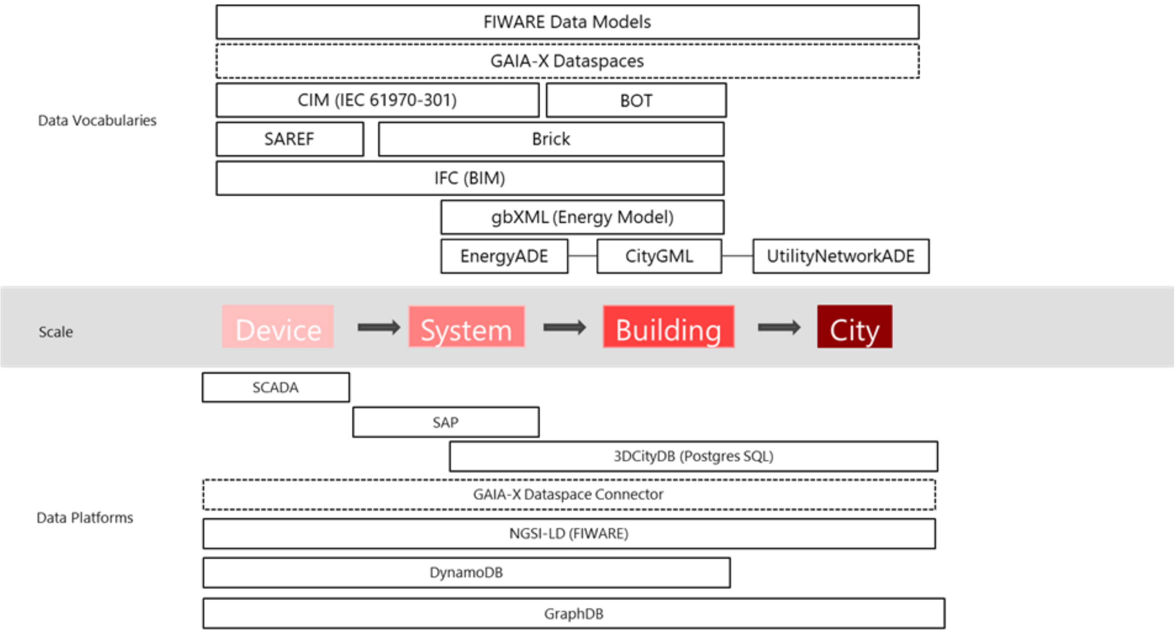


Figure 5: A summary of the reviewed data vocabularies and data platforms. The scale indicated the physical scale of the data they represent.

4 Approach

At the core of the platform is the DIGICITIES ontology, this provides the data structures to organise the semantic layers and ensure the requirements of each service are met. The development and structure of the ontology is described in Section 4.1.

Data processing was required at several stages in the project including: the conversion of data into the semantic layers, cleaning and processing of timeseries data for ML processes, the transformation of data to make it suitable for the use cases. These steps are detailed in Section 4.3. The architecture of the prototype platform is detailed in Section 4.4.

4.1 Collaboration

The project collaboration followed the tasks as responsibilities set out in the proposal. Here is a summary of the contributions from each partner involved in the project.

- **Empa** led the project and were responsible for the data structures, Section 4.2, and development of the platform architecture and modular tools detailed in Section 4.4. The developed ontology is planned to be released as an open-source repository. The current platform is hosted at https://platform.digicities.ch/. However, the cost of maintaining and hosting this platform as a cloud service needs to be considered for the future. Another possibility is to release the code open source so that it can be locally deployed and act as semantic tooling after the project completion; however, additional work is required to complete this task. This option moves away from the envisioned centralized data hub for energy data; however, it also provides a decentralized solution where the users are responsible for locally hosting their own workspace. This option is currently being considered in the REFORMERS EU Project where the tooling developed in DIGICITIES is adding significant value to the activities. If interested, please contact james.allan@empa.ch or george.mavromatidis@empa.ch for the final decision regarding platform deployment. Empa was also responsible for the scoping, execution and evaluation of the Swiss UC2 Energy Planning and Resource Optimisation detailed in Section 5.2.
- **SUPSI** led the data processing of ML algorithms for the hydro scheduling use case and were the main contributors of Section 4.3. SUPSI was also responsible for the scoping, execution and evaluation of Swiss UC1 Operational Decision Making detailed in Section 5.1.
- **Lugano Living Lab** coordinated with stakeholders in the region and organized events with the need-owners of the Swiss use cases.
- **AIT** was responsible for coordinating and implementing the project in Austria with their own use cases and stakeholders.
- **Microsoft** provided funding to work with one of their service partners on implementing Azure services within the project.
- **Sympheny** are service provider in the energy domain and assisted the development of third-party APIs in communication with the platform.
- **BeyondCivic** provided its data sharing technology stack (TrustRelay) to onboard organization and users, govern their data and perform trusted data sharing transactions based on the concept of data products.

Regular meetings throughout the project ensured that the developments remained aligned across collaboration partners.

4.2 Data Vocabularies

The DIGICITIES ontology was developed to accommodate the diverse services and models that require data from the semantic layers. The reviews of domain ontologies (3.5 Review of Data Vocabularies) recognised that there are many ontologies representing different parts of the energy domain; however, service developers still encounter issues when trying to apply a specific ontology or data model to their own use case or application. Therefore, rather than trying to define another domain ontology, the focus was placed on accommodating individual service needs by providing an internal flexible data model, the DIGICITIES Core Ontology. The ontology is written using the Web Ontology Language (OWL) to provide a framework for semantic reasoning. This core model focuses on defining consistent structures for attributes and then linking to them to components. It also provides a mechanism to link term with those in established domain ontologies. The role of the DIGICITIES core ontology is shown in Figure 6**Fehler! Verweisquelle konnte nicht gefunden werden.**.

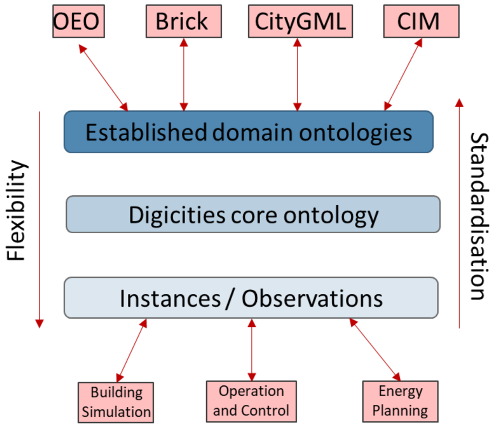


Figure 6: DIGICITIES core ontology

This section details the design of the ontology and its intended use.

4.2.1 Ontology Architecture

The DIGICITIES core ontology establishes a hierarchical structure organised around five primary component categories based on functional analysis of energy systems: Physical Infrastructure (devices, networks, storage), Process Components (conversion, transport, storage processes), Flow Components (energy carriers, materials, information), Spatial Components (locations, geographic context), and Actor Components (stakeholders, operators, regulators). This organisation provides semantic precision whilst maintaining flexibility for diverse energy applications.

The Device classification includes sensors, actuators, controllers, meters, and switches. Process hierarchies distinguish between conversion, transport, and storage processes, whilst Flow components differentiate between Energy Carrier Flows, Material Flows, and Information Flows. The hierarchical structure implements inheritance relationships where universal properties apply through the root Component class, whilst specialised properties address specific component categories.

4.2.2 Attribute Classification System

The attribute taxonomy implements a comprehensive classification system organised by fundamental data type rather than temporal behaviour. This approach recognises that temporal characteristics apply across attribute categories, enabling flexible modelling patterns where individual attributes can exhibit both time-invariant and time-varying properties.

Physical Attributes represent quantitative properties with defined measurement units, integrating with QUDT vocabularies through the hasDefaultUnit property. Cost Attributes address economic properties through SimpleCostAttribute (fixed values) and UnitBasedCostAttribute (unit-based pricing) subclasses, supporting currency specification through FIBO vocabularies. Curve Attributes represent functional relationships through coordinate pair collections with mandatory x-axis and y-axis unit specifications. Custom Physical Ratio Attributes support ratio-based measurements with distinct numerator and denominator units for derived quantities that are not included in QUDT.

Categorical Attributes handle discrete states through controlled vocabularies, Geospatial Attributes address location-based properties for spatial analysis, Event Attributes capture temporal occurrences with configurable precision (annual to timestamp resolution), and Simple Value Attributes represent unitless quantities such as configuration parameters.

4.2.3 Temporal Data Handling

The ontology addresses temporal representation through three distinct TimeSeries subclasses: HistoricTimeSeries (measured past data), LiveTimeSeries (real-time measurements), and FutureTimeSeries (forecasted projections). Dynamic attributes maintain simultaneous references to multiple temporal variants through specialised data properties, enabling comprehensive temporal analysis where individual attributes reference historical validation data, real-time monitoring streams, and future scenario projections within unified semantic structures. Common temporal metadata properties include temporal bounds (startTime, endTime), temporalResolution, and data location information.

4.2.4 Interoperability and Mapping

The ontology incorporates established vocabularies whilst implementing systematic mapping capabilities to support interoperability with domain standards. Integration with QUDT enables standardised unit representation and dimensional analysis, whilst FIBO integration ensures standardised currency handling. The mapping framework implements three relationship types: equivalence relationships (owl:equivalentClass, owl:equivalentProperty) for semantic identity, hierarchical relationships (rdfs:subClassOf, rdfs:subPropertyOf) for taxonomic connections, and similarity relationships (skos:closeMatch) for semantic proximity. This approach supports energy system models requiring simultaneous compliance with multiple domain standards, such as building energy models requiring SAREF alongside electrical grid models utilising CIM standards.

4.2.5 Scenario Management

The Scenario class provides the foundation for alternative system representation, with the derivedFrom property establishing hierarchical relationships between scenarios. Component inclusion operates through the usedInScenario property applicable to components, attributes, and component links, enabling different system configurations within unified knowledge bases. The ComponentLink class represents relationships between components that vary across scenarios, supporting different system topologies and operational patterns. The Assumption framework implements systematic scenario parameterisation through targetComponent, targetAttribute, and modifier properties, enabling scenario definition through assumption sets rather than complete model duplication.

4.2.6 Implementation Process

The knowledge graph construction comprises two main components: classes and attributes collected via standardised Excel templates specifying component names, attributes, and semantic structure information; and physical relations handled through the linksComponent object property and subproperties linking physically connected infrastructure. Service data requirements are defined using YAML specifications that reference DIGICITIES components and attributes, maintaining semantic consistency with formal ontology definitions whilst providing practical specification formats validated against domain and range restrictions.

4.3 Data Processing

4.3.1 Cleaning and Processing of Data

To ensure the quality and usability of the data for analysis and modeling, we applied a series of cleaning and processing steps. These included:

- **Data distribution and consistency checks.**
  We examined the characteristics, time evolution, and statistical distribution of each variable and verified consistency with prior knowledge and mechanistic models, identifying unexpected patterns or anomalies.

- **Outlier removal and smoothing.**
  Outliers were detected and removed. For time series data, smoothing techniques such as moving averages were applied to highlight underlying trends.

  Missing data analysis and imputation.
  We assessed the amount and distribution of missing data to identify non-random patterns. For tabular data, we found Multiple Imputation by Chained Equations (MICE) very effective. For time series, classical methods such as rolling mean and interpolation were preferred. When sufficient training data was available, we found that using machine learning models (LightGBM) to predict and impute missing value can capture complex patterns, enabling more accurate imputations. Finally, we also found that for modeling, using machine learning methods that can handle missingness without imputation, such as LighGBM, is a valid alternative.

- **Units conversion.** All measurements must be standardized to consistent units across datasets to ensure comparability and avoid misinterpretation.
- **Feature engineering and derivation.** New features can be derived both from prior mechanistic knowledge and from transformation of the raw signals to enrich the dataset and improve model performance. For time-series data, this includes aggregations and transformations such as lagged variables, rolling-window statistics (e.g., local maxima, minima, means, and other window-based summaries).

- **Temporal alignment and normalization.** Time series with different time steps, offsets, or scales must be aligned and normalized to enable integration and comparison across sources, which is particularly important in energy-related applications.
- **Predictive Modelling Preparation**. The cleaned and processed data must be structured for use in time series predictive modelling, including formatting, feature construction and selection.

4.3.2 Use case specific data transformation

For the signals used in Use Case 1 – HPP Scheduling, we manually supervised the data and its distribution, removing only a few outliers. During this process, we identified a frequent inconsistency between the measured flow rate and the reconstructed flow rate, which was derived from production data and reservoir level. We therefore chose to use the reconstructed flow rate, defined as:

(Change in reservoir volume per unit time) – (Water consumption at the given production level per unit time)

Reservoir volume estimates were derived from observed water levels using an interpolated function based on the level–volume characteristic curve supplied by the utilities.

We also analyzed missing data, which in some cases corresponded to extended periods. Imputation was performed using traditional methods (e.g., linear interpolation) only when a small number of consecutive values were missing, and exclusively for simulation purposes in scheduling. No imputed data were used for model training, as we employed LightGBM, which natively handles missing values.

To enrich the dataset, we integrated additional sources of information, specifically weather forecasts from OpenMeteo. These included forecasts for the utility service area which were used in demand modeling and forecasts for the basin catchment area which were used in flow rate prediction.

When using weather forecasts, it is important to align them with the demand prediction horizon. For instance, if the model predicts demand 24 hours ahead, the corresponding weather forecast must be one that was issued at least 24 hours before the target time — in other words, the forecast must reflect the information that would have been available at the moment the 24-hour-ahead prediction is made.

It is important to note that historical data on weather forecasts often only include the latest forecast produced before the target time, that is, typically 12 hours earlier. As a result, models were trained (and sometimes tested) using as inputs weather forecasts produced at most 12 hours before the target time but later applied using longer-term predictions (up to 48 hours ahead), which are inherently less accurate. This mismatch is not optimal, as it may lead the model to overestimate the importance of weather features, and the estimated performance may be overly optimistic. However, when tested with realistic forecast horizons, the models still performed well. Where possible, it is recommended to retain all available weather forecasts, to enable more accurate and horizon-consistent training and evaluation. (Change in reservoir volume per unit time) – (Water consumption at the given production level per unit time)

Reservoir volume estimates were derived from observed water levels using an interpolated function based on the level–volume characteristic curve supplied by SES.

We also analyzed missing data, which in some cases corresponded to extended periods. Imputation was performed using traditional methods (e.g., linear interpolation) only when a small number of consecutive values were missing, and exclusively for simulation purposes in scheduling. No imputed data were used for model training, as we employed LightGBM, which natively handles missing values.

To enrich the dataset, we integrated additional sources of information, specifically weather forecasts from Open Meteo. These included forecasts for the utility service area which were used in demand modeling; forecasts for the basin catchment area which were used in flow rate prediction.

Including weather forecasts requires a careful alignment of weather forecasts with demand forecasts, based on the prediction horizon. For example, to train a model that predicts demand 24 hours ahead, the weather forecast used must have been produced at least 24 hours before the target time.

It is important to note that historical data often only include the latest forecast available before the target time, typically with a maximum horizon of 12 hours. As a result, models were trained (and sometimes tested) using 12-hour ahead forecasts but later applied to longer-term predictions (up to 48 hours ahead), which are inherently less accurate. This mismatch is not optimal, as it may lead the model to overestimate the importance of weather features, and the estimated performance may be overly optimistic. However, when tested with realistic forecast horizons, the models still performed well. Where possible, it is recommended to retain all available weather forecasts, to enable more accurate and horizon-consistent training and evaluation.

4.4 Architecture and Interfaces

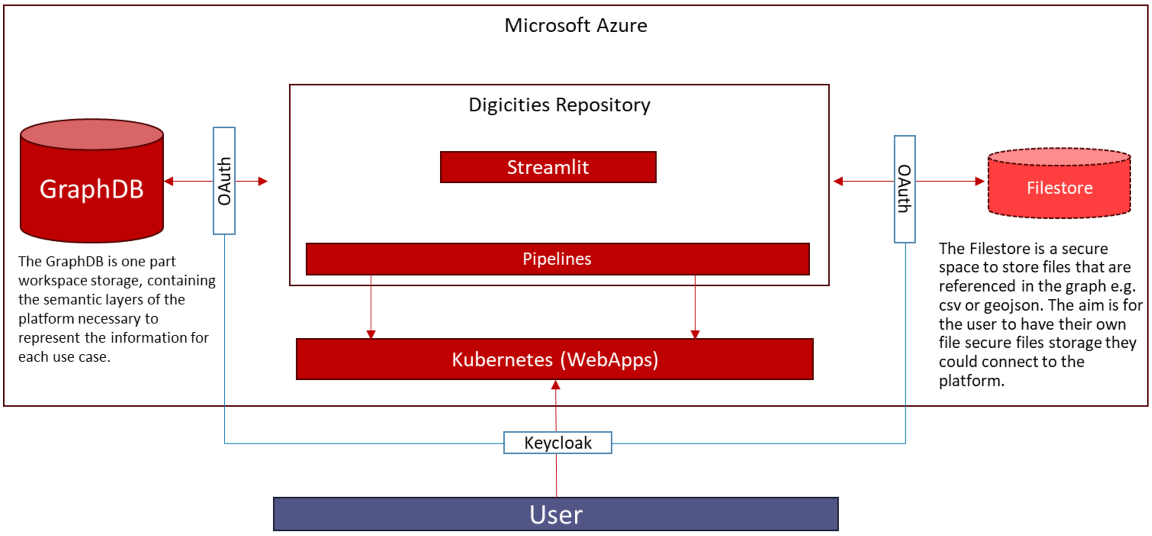
The platform is an interface to work with the semantic layers. This section describes the architecture adopted in the prototype platform. The primary functionality is a secure environment that allows users to access their data; however, the platform also contains a set of modular functions to allow the users to work with the semantic layers. This section describes the core architecture of the platform, the different modules and the interface with third party applications.

4.4.1 Core Architecture

The core architecture of the DIGICITIES platform is comprised of the following components:

- Microsoft Azure tenant
- Frontend prototype (Streamlit)
- RDF Triplestore containing the semantic layers (Ontotext GraphDB)
- Cloud storage (NextCloud)
- User identity and access management (Keycloak)

The organisation and interaction of these components is shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**.



The GraphDB is one part workspace storage, containing the semantic layers of the platform necessary to represent the information for each use case.

The Filestore is a secure space to store files that are referenced in the graph e.g. csv or geojson. The aim is for the user to have their own file secure files storage they could connect to the platform.

Figure 7: Core architecture of the DIGICITIES platform.

The following section gives an overview of the submodules of the user interface of the platform and their functionality. The development of the Streamlit user-interface was assisted by Claude.ai (Anthropic, US).

4.4.2 Ontology Manager

The Ontology Manager provides a comprehensive platform for building, extending, and maintaining the DIGICITIES core ontology, abstracting the complexities of RDF/OWL syntax whilst enabling semantic model development. The system maintains clear separation between the core ontology and user-generated extensions, storing these separately whilst providing seamless merging capabilities for deployment.

4.4.3 Component and Attribute Management

Users can define hierarchical component structures by specifying labels and parent classes, with the system automatically managing inheritance chains and creating associated attribute categories. Components can be reorganised through parent relationship modifications, with the platform updating all dependent structures. The platform supports the multiple attribute types detailed previously—Physical, Cost (Simple and Unit-Based), Curve, Categorical, Geospatial, Custom Physical Ratio, Event, and Simple Value—each with specialised properties and validation rules. When attributes are linked to components, the system automatically creates necessary property definitions and ensures proper categorisation within component attribute hierarchies, handling the complex RDF graph structure behind simple selection interfaces.

4.4.4 Mapping and Interoperability

The Ontology Manager provides comprehensive mapping capabilities for external ontology integration. Users can upload external ontology files and create alignments between DIGICITIES concepts and external vocabularies, supporting the equivalence, hierarchical, and similarity relationship types described in the interoperability framework. These mappings are stored separately and managed independently, allowing simultaneous connections with multiple external standards. The interface presents internal and external concepts side-by-side, enabling informed mapping decisions.

4.4.5 Deployment and Exploration

Query-based interfaces enable users to explore components, attributes, and properties throughout the development process, presenting information in structured tables that show inheritance relationships, labels, and metadata. Once complete, ontology extensions can be deployed directly to GraphDB triple store repositories, with the module handling authentication, serialisation with proper namespace bindings, and upload to named graphs, enabling immediate support for data ingestion and querying operations across the platform.

4.5 Replica Builder

The Replica Builder enables users to create digital replica that conforms to the ontology structure defined in the Ontology Manager. This module creates instances that conform to the ontology. These instances are digital replicas of real-world components and their relationships, generating RDF graphs that populate the workspace.

4.5.1 Instance Creation Workflow

The module begins by loading the ontology from GraphDB, automatically discovering available component types and their permissible attributes. Users create component instances by selecting a component type and providing a unique identifier and label. The system generates URIs based on configurable patterns and organizes instances by type for easy navigation. Instance management includes search and filtering capabilities, with automatic validation to prevent duplicate identifiers.

4.5.2 Attribute Configuration

Users configure instance attributes through a type-specific interface that supports multiple attribute types including physical measurements with units, dynamic time series references, categorical values, temporal events, costs, curves, resource paths, and annotations. The system constrains attribute selection to only those defined in the ontology for each component type, ensuring semantic validity. Input fields adapt dynamically based on attribute type, presenting appropriate validation and data entry options.

4.5.3 Relationship Management

The Link Manager enables users to define relationships between instances using properties from the ontology. Users select source and target instances and choose from available relationship properties such as spatial containment, connectivity, or supply relationships. All links are validated against ontology constraints and stored in the system description graph.

4.5.4 Data Import and Deployment

For bulk data loading, the module provides Excel import functionality that processes multi-sheet workbooks where each sheet represents a component type. Users can preview generated RDF, validate against ontology constraints, and choose to merge with or replace existing data. The deployment interface offers two modes: append mode adds new instances to existing graphs, while replace mode clears and rebuilds graphs. Users can preview generated Turtle serialization, download for backup, and upload directly to GraphDB repositories with automatic namespace binding and transaction management.

4.6 Replica Explorer

The Replica Explorer provides an interactive interface for browsing and analyzing component instances stored in the knowledge graph. The module automatically discovers all component types with instances, presenting them in a dropdown menu with instance counts. Users select a component type to view comprehensive tables displaying instances and their attributes, with support for multiple attribute types including physical measurements with units, costs with currencies, categorical values, temporal data, time series references, curve data points, geospatial information, and custom unit ratios. The interface provides search and filtering capabilities, automatically formats values for readability, and enables data export as CSV files. For curve attributes, users can generate interactive visualizations with statistical summaries.

4.7 Query Manager

The Query Manager provides an interface for creating, editing, executing, and managing SPARQL queries to interrogate the knowledge graph. Users can create new queries or select from saved workspace-specific queries stored in global Nextcloud storage, with default query templates provided based on workspace type (e.g., Renewable Energy, Municipal Infrastructure, Building Management). The query editor allows users to write and modify SPARQL queries with syntax highlighting, then execute them against the connected GraphDB repository. Results are displayed in interactive tables with configurable namespace replacement and path extraction options to improve readability by converting long URIs into shortened prefixes or local names. Users can save, update, or delete queries with overwrite protection, and export results as CSV files in either formatted or raw format. The module maintains workspace-specific query collections and namespace configurations, enabling users to build libraries of reusable queries for common data exploration and analysis tasks across their digital twin knowledge base.

4.8 Data Products

The Data Products module enables users to load, explore, and visualize self-contained data packages that combine semantic TTL descriptions with associated resource files. Data products are organized as folders containing a TTL file defining component instances and a resources subfolder with related data files such as time series CSVs, GeoJSON geometries, or weather data.

4.8.1 Loading and Management

Users can load data products from two locations: open data products available globally to all users, and private data products specific to their workspace. The module connects to Nextcloud storage using environment-based authentication, automatically discovering available data product folders in both locations. When loaded, the system parses the TTL file to extract component definitions and their attributes, then lists available resources in the resources subfolder. Users see summary metrics showing total products loaded, component counts, and available resource files, with the ability to select specific products for detailed exploration.

4.8.2 Data Exploration

The explorer interface presents data products with expandable component listings organized by type. Users can filter components and view detailed information including URIs, labels, attribute counts, and linked resources. Attributes are grouped by category (physical, cost, geospatial, dynamic, categorical, etc.) and displayed in tables with formatted values, units, and resource indicators. The interface handles multiple attribute types with appropriate formatting, converting URIs to readable values and truncating long strings. Users can access raw JSON views of attribute data and download component information for external analysis.

4.8.3 Visualization and Analytics

The module automatically determines appropriate visualization types based on component attributes. For geospatial data with latitude/longitude coordinates, it renders interactive maps showing component locations. Time series attributes trigger line charts with range sliders and statistical summaries showing min, max, mean, and data point counts. Curve data is visualized with scatter plots and data point tables. Components without specialized data types display static attribute tables. The visualizer loads resource files on demand, parsing CSVs into dataframes, processing GeoJSON for mapping, and handling other file formats. Users can download visualizations and underlying data, navigate between components, and switch between visualization and detailed attribute views to explore data from different perspectives.

4.9 Service Registration

The Service Registration module enables users to create YAML configuration files that define service requirements by mapping ontology components and attributes to service input fields. Users connect to GraphDB to load available components and their valid attribute mappings from the dici_onto ontology, ensuring only semantically correct configurations can be created. The interface supports building hierarchical component structures with parent-child relationships, where users select component types and organize them using automatically generated link patterns. For each component, users configure which attributes to include and designate them as Static, Historic, Live, or Future data types, with multiple types selectable simultaneously. A validation system checks all component-attribute pairs against ontology mappings, displaying errors, warnings, and valid configurations with actionable suggestions for fixing issues. Users can customize YAML field names while maintaining ontology references, preview the generated structure, and export configurations as YAML files for service integration and automated data extraction workflows.

4.10 Scenario Builder

The Scenario Builder enables users to construct scenario configurations by selecting component instances from data products and knowledge graphs that satisfy the data requirements specified in service YAML files created through the Service Registration module. This module bridges service definitions with available instance data, allowing users to create multiple scenario variants of the same system for different analysis contexts.

4.10.1 Service-Driven Configuration

Users select a service definition YAML file that specifies required component types, attributes, and relationships. The system parses the YAML to extract component requirements including hierarchical link patterns, required attributes with nested properties and time series references, and temporal data specifications. The interface displays which component types are needed and validates whether selected instances provide all required attributes, ensuring scenario completeness before export.

4.10.2 Multi-Source Instance Selection

The builder integrates component instances from multiple sources: the workspace knowledge graph containing custom components, private TTL data products specific to the workspace, and global TTL data products shared across all workspaces. Users enable desired data sources and select component instances through a batch selection interface, with the system tracking data provenance and validating each instance against service requirements. This allows flexible scenario construction even when the workspace knowledge graph is empty, relying entirely on data products if needed.

4.10.3 Relationship Definition and Validation

Users define relationships between selected components according to link patterns specified in the service requirements, such as parent-child hierarchies or connectivity relationships. The validation system ensures all required attributes are present in selected instances, checks nested property requirements for complex attributes, and verifies that component relationships match service specifications. Components are color-coded by validation status, showing which instances fully satisfy requirements and which have missing attributes.

4.10.4 Scenario Serialization and Variants

The system generates scenario-specific TTL files containing the selected component instances, their configured attributes and relationships, and service-specific metadata. These scenario graphs exist as separate entities alongside the original instance data, enabling users to create multiple scenario variants of the same system for different analytical purposes, temporal contexts, or configuration alternatives. Each scenario is serialized as an independent TTL file that can be uploaded to the knowledge graph, NextCloud or downloaded for conversion for external services, maintaining clear separation between base instance data and scenario-specific configurations.

4.11 Assumptions

The Assumptions module enables users to create scenario variants by systematically modifying component attributes in baseline scenarios through predefined assumptions or manual edits. This allows exploration of different system configurations, temporal changes, or what-if analyses while maintaining clear provenance tracking from baseline to modified scenarios.

4.11.1 Baseline Loading

Users load baseline scenarios from the workspace scenarios folder (created by Scenario Builder) or upload TTL files directly. The system parses baseline scenarios to extract components, attributes, and relationships, displaying summary metrics showing component counts, types, and total attributes. Loaded baselines serve as the foundation for generating modified scenario variants through assumption application or manual modification.

4.11.2 Predefined Assumption Application

Users select from predefined assumptions that target specific component types and attribute categories such as physical properties, costs, or geospatial data. Assumptions can be single modifications creating one variant or series modifications generating multiple scenarios across timesteps for temporal analysis. The system validates assumption compatibility against baseline components, showing which assumptions can be applied and how many components will be affected. Applied assumptions automatically modify relevant attribute values according to specified modifiers like percentage changes, absolute adjustments, or scaling factors.

4.11.3 Manual Modification

Users can manually edit individual component attributes through an interactive interface, selecting components by type or search term and modifying attribute values directly. The system tracks pending modifications with before/after comparisons and percentage changes, allowing users to review and adjust edits before generating scenarios. Manual modifications support all attribute types including physical measurements, costs, categorical values, and geospatial data, with appropriate input controls for each type.

4.11.4 Scenario Generation and Export

Modified scenarios are generated as independent TTL files with full traceability, including metadata about applied assumptions, modification types, and affected component counts. Users can view generated scenarios with modification logs showing which attributes changed and by how much, preview TTL content, and export scenarios individually, as ZIP archives, or upload directly to the workspace scenarios folder. Each generated scenario exists as a separate entity that can be used independently for analysis, uploaded to the knowledge graph, or serve as a baseline for further modifications, enabling iterative scenario exploration and variant creation.

4.12 API Submission

The API Submission module enables users to convert scenario TTL files into service-specific JSON/YAML formats and submit them to configured external APIs for analysis, simulation, or optimization. The module handles the complete workflow from TTL conversion through submission to result management, with automatic result saving to NextCloud for persistence and review.

4.12.1 Service Configuration and Templates

Users configure API endpoints by selecting from service templates loaded from NextCloud or local files, specifying connection details including URLs, HTTP methods, authentication credentials, and timeout settings. The system supports multiple authentication types including bearer tokens, API keys, basic authentication, and environment-based JWT tokens. For the pilot we configured a fast set up for CESARP building simulation and Sympheny regional analysis, automatically configuring endpoints with appropriate defaults while allowing customization of polling intervals, retry limits, and authentication parameters.

4.12.2 TTL to JSON/YAML Conversion

Users select scenario TTL files from the workspace scenarios folder or upload files manually, then choose target service templates for conversion. The conversion engine parses TTL content to extract component instances, attributes, and relationships, mapping them to service-specific data structures according to template specifications. The system handles complex hierarchical component links, nested property references, time series attributes, and categorical values, generating properly formatted JSON or YAML output that matches each service's expected input structure while preserving all semantic relationships and data provenance.

4.12.3 Batch Submission and Polling

Users select multiple converted scenarios for batch submission to configured APIs, with the system handling sequential processing and progress tracking. For services requiring asynchronous processing like Sympheny, the module automatically manages job submission and status polling, displaying real-time updates on job progress and completion. Submissions include configurable retry logic, timeout handling, and error recovery, with detailed logging of request/response cycles and status codes for troubleshooting failed submissions.

4.12.4 Result Management and Persistence

All submission results are automatically saved to NextCloud in structured JSON format, organized by service name with timestamps and metadata including submission status, API responses, scenario IDs, and result URLs. Users can browse past results through a dedicated viewer showing submission history, success rates, and detailed response data. For Sympheny submissions, the system captures and displays multiple result URLs including scenario views, execution file locations, and dashboard links, enabling direct navigation to external analysis platforms. Results remain accessible across sessions and can be downloaded individually or in bulk for offline analysis, reporting, or integration with other tools.

5 Description of use cases

A key objective of the project is to meet the needs of stakeholders in the energy sector. A series of use cases have been defined to address needs identified in surveys and discussions with stakeholders. Each use case is unique but has similar requirements when managing the data of their digital replica. Two distinct use cases were defined through project workshops and stakeholder communication. Each use case is summarised below with a list of the objectives and metrics to be considered during the measurement and evaluation phase.

5.1 UC1 Operational Decision Making

This use case focuses on optimising energy storage management by assisting operational decision-making. In particular, we focus on water storage for small hydro power plants and battery storage for self-consumption communities. Utilities and self-consumption communities often rely on them to shave the energy demand peaks. This may reduce the cost of energy, as in self-consumption communities better matching between the consumption and generation can maximize self-consumption, while for utilities, smaller peaks correspond to a lower price on the energy market. On the other hand, a smoother profile of the energy demand facilitates the distribution grid operation.

Optimal management of energy storage is often based on predictive control, in which decision-making is based on energy demand forecasts. Those forecasts require managing different sources of data, manipulating, combining, and constantly updating them to train data-driven forecasts of the relevant time series (energy demand, photovoltaic production, availability of water, etc.). Besides, high quality data must be fed into the controller in real-time to allow its field operation.

Data scientists recognize the data pre-processing step as the most time-consuming and least enjoyable step[3]. Thus, although utilities acquire the skills necessary to implement machine learning approaches, the difficulties of obtaining reliable data and managing and cleaning them often represent the greatest limitation to implementing forecasting models that may help energy management.

To meet this need, DIGICITIES is building an infrastructure that will organize the data to be readily available for training and operating machine learning models, focusing on time series forecasts, which are involved in a large fraction of ML applications in the energy field.

5.1.1 UC1 Objectives

- Evaluate the time savings made possible by DIGICITIES in developing time series predictive models and improve flexibility. This will be compared to developing it from original data managed by the companies and eventually combined with public repositories (e.g., for weather forecasts or measurements).
- Show how such models allow improved peak shaving by optimal control for current operator scheduling, resulting in a consequent reduction of the cost of electricity and optimization of water usage for clean energy production.
- Integrating battery control, the data-driven predictive model can improve community self-consumption and reduce demand peaks at different grid levels.
- Establish through simulations how battery capacity influences peak reduction at different NE levels.

---

[3] Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. Forbes Inc. https://www.forbes.com/sites/gilpress/2016/03/23/data -preparation-most-time-consuming-least-enjoyable-datascience-task-survey-says/#4db1b036f637, 2016.

5.1.2 UC1 Measurement and Evaluation

The KPIs for this use case will be compared against a baseline control strategy implemented in the power plant. The performance achieved by the optimal scheduling will be assessed using simulations with historical data comparing the effectiveness of the operator decisions with those made by the controller based on the observations (unrealistic, best-case scenario) or simple ML models (baseline scenario). The following KPIs will be compared between the baseline and the DIGICITIES solution.

Impact Category	**KPI**
Data Access	Number and size of time series used for training the models
	Number and quality of new data available (forecasts)
	Time for training data preparation and validation set up
Decision making	Comparison of historical and simulated decision
	Comparison of operator and controller decisions
	Support provided in the choice of the battery size
Energy Impacts	Peak reduction
	Energy cost reduction due to the reduced size of monthly peaks
	Increased self-consumption of the self-consumption community exploiting optimized battery management

5.1.3 Machine Learning for Operation and Control

The objective of this use case is leveraging energy storage to smooth out the load curve and maximize self-consumption by optimal usage of water storage resources in HPPs and battery charge in small self-consumption communities during peak demand periods. This requires addressing peak energy demand while optimizing available resources. Local HPPs dedicated to peak load shaving rely heavily on the expertise of operators to meet this challenge effectively, while batteries are managed by charging them when PV production exceeds community demand and discharging when demand exceeds PV production.

Concerning HPP production scheduling, model predictive control (MPC) is used to manage resources in the small-size HPP managed by the AEM, a utility distributing energy primarily to the municipalities of Massagno, Capriasca and Isone in Ticino, Switzerland and in the HPP of Società Elettrica Sopracenerina SA (SES), which operates in 38 municipalities of the Canton of Ticino, in the Sopraceneri region. The price of the electrical energy paid to the cantonal electricity supplier (AET - Azienda Elettrica Ticinese) is based on both the energy volume and the power transfer price, which is proportional to the peak demand reached each month. AEM and SES exploit their HPP primarily to shave peak demand, thus reducing the grid costs.

AEM plant collects water from two rivers and accumulates it in a surge tank from which it can draw to produce during the peaks, as described in **Fehler! Verweisquelle konnte nicht gefunden werden.**. The maximum production of the power plant is 4 MW, achieved at a water flow rate of 2.11 m^3/s. The reservoir can store about 18,500 m^3 of usable water, based on the specified minimum and maximum reservoir water levels, corresponding to approximately 9.75 MWh of stored energy.

SES collects the water in the Vasasca dam reservoir, which can store about 441,000 m^3 of usable water corresponding to about 360 MWh of stored energy. The maximum production of the power plant is 9 MW.

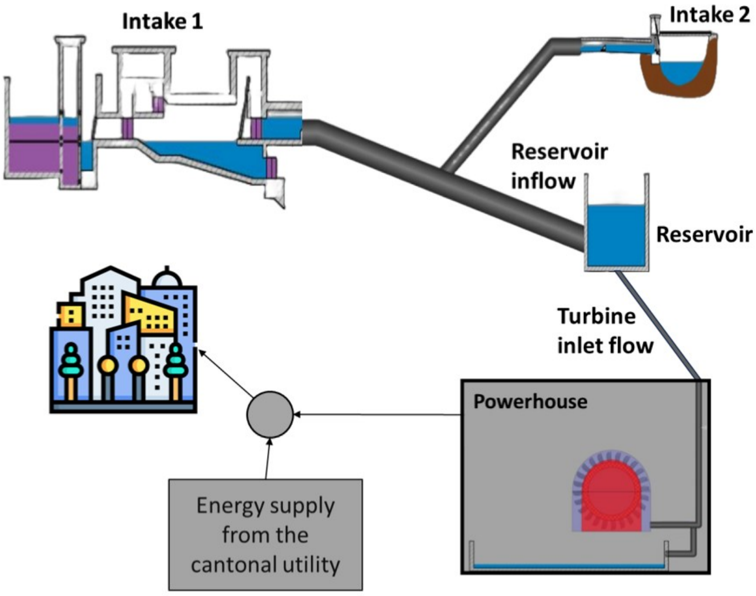


Figure 8: Scheme of AEM hydro power plant.

The production scheduling of the hydroelectric plants is a combination of real-time corrective control (that adjusts system operations in response to observed variations in demand) and operator expertise. The operators' primary objective is to minimise the maximum daily peak of electricity paid to the provider AET (usually observed in the evening). When the reservoir (surge tank) is full, the plant is set to reduce the monthly peak as much as possible. If the storage reservoir is not full, the operators must still try to reduce the current peak and ensure that sufficient water levels in the reservoir will be available in future. The operators determine the target value of the daily peak based on their expertise. Reactive control then schedules the production to meet the difference between the demand, assumed equal to the last observation available, and the target peak value. Operators can also manually intervene to schedule the production when weather conditions change. For instance, if the external temperature drops significantly in winter, they will increase production, foreseeing an increased demand.

Thanks to their many years of experience, operators can ensure efficient utilization of resources and contribute to a stable energy supply. However, their work, and even more so for less experienced operators, can be supported and decision-making and operational efficiency enhanced by an automatic system that provides them with accurate forecasts of short-term energy demand, water availability, and an optimal automatic scheduling system that addresses the utility's need to minimize current and future peak demands.

In a previous study (Derboni et al., 2021), IDSIA researchers applied a methodology called Model Predictive Control (MPC) to the flexible scheduling of residential energy loads. This approach leveraged measurements from smart meters for non-intrusive load monitoring. The scheduling component primary function was to exploit controllable flexibilities for reducing the peak in the energy demand of a self-consumption community. The scheduler considered control actions over a future period and based its decisions on demand estimates produced by a time series predictive model. This model hybridizes a seasonal method with a simple classification model. Power demand is assumed to have daily seasonal patterns, and future demand is thus derived as the average demand observed at the same time of the day over a given window of several days in the past. Observations are classified into five categories to account for the weather conditions, which may affect the demand based on the observed temperature and irradiation conditions. Only observations in the same weather category are averaged to predict the demand at a given future time. We use this approach as a comparison baseline for our new predictor.

Our work aims to create a forecaster of the energy demand based on historical data from the utilities. Such a predictor serves a dual purpose: it informs the scheduler’s decisions and provides valuable insights to the operator. We based such a predictor on lightGBM. A boosted tree-based algorithm which has proven very effective and popular for time-series forecasting because, thanks to its versatility, effectiveness, and relatively simple usage, it allows obtaining models with competitive accuracy in relatively limited development times and computational resources. For instance, boosted trees algorithms were one of the most used frameworks adopted by teams in the BigDEAL Challenge 2022, which was about energy load and peak forecasting. However, lightGBM does not provide tools for multi-output regression. Traditional approaches to multi-output predictions are based on training multiple independent or chained models. Some recent works have explored extensions of the gradient-boosted trees framework to handle multiple outputs. In our application, we must predict energy demand over a two-day-ahead horizon with a 5-minute resolution. Thus, our predictor must generate hundreds of output variables (precisely 576), making it preferable not to use a different predictor for each output. We propose a forecaster trained using LightGBM, which addresses the lack of native multivariate regression support by a straightforward yet effective strategy that exploits LightGBM's capability of handling missing data to learn a global model that is sufficiently accurate across the entire considered time horizon.

Historical data provided by the utility are used to train the energy demand forecaster and compare it to the baseline statistical model used by Derboni et al (2021). These experiments allowed us to evaluate the impact of the improved energy demand forecasts on the subsequent decisions made by the scheduler and compare these decisions with those made by operators. These comparisons highlight the potential of these methods to support operators in their daily decision-making processes and thus enhance energy storage management.

MPC is also used to control the charging and discharging of the battery that stores excess PV production in the Lugaggia Innovation Community (LIC). The demand and generation forecasts are processed by the controller, which optimizes the use of the stored energy to maximize self-consumption.

5.1.4 Power production scheduling and energy demand forecasting

The production of the HPP is controlled using an MPC framework. A similar approach is used for the control of the battery. MPC is a control scheme representing the system being controlled over a finite time horizon. Decisions are optimized along a control trajectory of 48 hours with 5-minute resolution. Only the decision for the first time slot is actuated, while the following ones are revised at the next time slot based on updated measures of energy demand, water availability, and weather. The HPP we model in the MPC framework is a plant without pumping facilities, i.e., the water storage is filled by inflow rivers only. This type of storage system is characterized by an uncontrollable inflow rate that can fill up to a given amount (when the storage is full the extra water spills out). We control the discharging power of the HPP at every timeslot via continuous variables pd_t.The scheduler uses the energy demand predicted at the current time t_0 for the (future) time t and select the sequence of pd_t values that minimize the maximum power requested from the grid during the projected horizon.

At each time slot, the energy demand predictive model produces 576 forecasts of the future energy demand using the following information:

- Past energy demand features include short-term lags (from 5 to 30 minutes behind with a step of 5 minutes) and medium-term lags going back of 1, 2 and 3 days and one week
- Calendar features include the categorical variables day of the week, month and a binary feature indicating if it is holiday in Ticino and, as numerical variables, the sine and cosine of the day of the year and the time of the day (to model circularity).
- Weather forecasts include an indicator of day and night hours, temperature and perceived temperature, precipitations, global horizontal radiation, global normalized irradiance and diffuse radiation, the latter three being supplied as instantaneous measurements and averaged over the previous hour.

Given that LightGBM does not inherently support multi-output regression and considering that, due to the high-dimensional nature of our output, we opted against constructing separate models, we designed a strategy that learns only two models: the first model provides predictions for the first time slot; the second model, instead, can be used to provide forecasts for whatever other step ahead h, from 2 to 576. We specifically train a model for the first time slot because this forecast is significantly more important than the others as it determines the only decision of the implemented scheduler. For more accurate forecasts, this first model uses the global irradiance, temperature and precipitation observed at the time t_0 instead of the weather forecast variables. For the second model, predictions for different values of h are obtained by adapting the inputs. For instance, if h corresponds to 24 hours ahead forecasts, we would supply as inputs the demand observed 1,2,3, and 7 days earlier, that is at the time of the prediction, t_0, and at $t_0 - 24h$, $t_0 - 48$, and $t_0 - 144h$ and the last weather forecast obtained before t_0 predicting the weather for $t_0 + 24h$. The short-term lags, instead, would be left missing, as they refer to future times for which observation are not available at the moment of the prediction, thus exploiting the ability of LightGBM to deal with missing values.

More details on the controller and forecasters are given in our paper Mangili et al. (2026).

5.2 UC2 Energy Planning and Resource Optimisation

The management of digital replicas using DIGICITIES aims to empower municipalities and utilities to generate and track the assumptions behind mandatory energy reporting and planning. Municipal planners require more control and awareness of the data used for reporting and planning in accordance with the objectives of the Swiss Energy Strategy. They have also expressed a need to minimize the uncertainty of forecasts and require a transparent process that can be updated and extendable to new data sources and assumptions, e.g., the inclusion of charging stations for electric vehicles. Once the sources of generation and consumption are established, they also need to take measures to optimize their operation.

A core objective of DIGICITIES is to connect digital services along the digital value chain. Each use case has its own value chain, and in this example, we have included a digital service provider that provides optimal energy planning and reporting based on data inputs. Our surveys revealed that digital service providers struggle with data preparation, resulting in additional features to be developed into their interfaces. This overcomplicates the process. The benefit of DIGICITIES is that it handles the digital replica to interface with the services offered by the provider efficiently. In this use case, we have partnered with a digital service provider to streamline the outputs of energy reports for energy system optimization. This use case has integrated two sub-cases – an energy planning module and an energy optimisation module.

5.2.1 Objectives

- Support obligatory municipal energy reporting
- Track assumptions and uncertainty associated with scenarios
- Provide energy planners with the best available data for their project
- Transparency of models used in forecasting and the quantification of uncertainty wherever possible
- Ability to track the progress and validity of a pathway
- Provide a means to evolve according to new datasets

5.2.2 Measurement and Evaluation

The baseline case is the process and result of the procedure used by the City of Lugano to create its existing strategic energy report. Each KPI will be evaluated before and after the implementation during the piloting phase. An overview of the KPIs to be considered are shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**.

Table 2: The KPIs to be evaluated in the Energy Planning and Resource Optimisation use case

Impact Category	**KPI**
Data Access	Quantity of data sources used in the energy planning study (comparison with existing study)
	Time saved accessing and processing data (estimate)
	Time saved creating usage agreements between partners (estimate)
Decision making	Accuracy of projected baseline (comparison with historic data)
	User evaluation
	Stakeholder reach
	Metrics reported in the energy planning report
Energy Impacts	Percentage of self-consumption achieved
	Carbon intensity of future scenarios

6 Results

The two use cases are described and evaluated in 6.1 Application and Evaluation and the different stakeholder groups are assessed in 6.2 Stakeholder Assessment.

6.1 Application and Evaluation

6.1.1 Use Case 1 – Model Predictive Control for Renewable Energy Resource Optimization

In this section, we describe how we applied a model predictive control (MPC) approach to optimize the use of renewable energy resources in response to grid demand. The controller, developed as described in Section 5.1.4, was designed to manage the use of energy reserves based on demand forecasts, with the objective of minimizing peak loads or maximizing self-consumption. We tested this approach in two distinct scenarios: (1) the management of small hydroelectric plants for peak shaving, and (2) the control of a battery system for storing solar energy produced by a small energy-sharing community.

Scheduling of HPP production. The first application addressed the needs of two utilities in Ticino (AEM and SES), which operate hydroelectric plants used to reduce peak demand during high-consumption hours. This is particularly relevant because the Swiss national transmission tariff (Swissgrid) consists primarily of a working tariff, charged in cents per kWh based on measured end-user consumption, and a power tariff, which is calculated from the average monthly peak load. The power tariff generates costs that increase linearly with the magnitude of the monthly peak, according to the relation: cost=(power tariff)×(monthly peak). According to the tariffs and remuneration rates for the transmission grid published by Swissgrid (https://www.swissgrid.ch/en/home/customers/topics/tariffs.html, accessed: 6 February 2026), the power tariff in kCHF/MW per year was 46.38 in 2024, 47.22 in 2025, and is 74.28 in 2026.

The goal of the controller in this context was to minimize the monthly peak by strategically dispatching hydroelectric power.

We applied the strategy described in Section 3.11 to the control of both AEM and SEM hydropower plants, using an identical structure for the control component and a very similar setup for the demand forecasting part.

Section 3.11 describes the modelling approach for AEM forecasters. For forecasting the energy demand of the SES utility, we adopt a similar approach, with adjustments reflecting the characteristics of the data. The forecasting horizon remains 48 hours, but the data resolution is 15 minutes, resulting in 192 prediction steps ahead. The same set of calendar features is employed, to capture daily, weekly, and yearly seasonal patterns. The past energy demand features are adapted to the coarser temporal resolution and include short-term lags of 15, 30, 45, and 60 minutes, as well as longer-term lags of 1, 2, and 7 days. For the long-term lags, we also include the minimum, maximum, and mean demand computed over 30-minute and 60-minute windows centered around the corresponding lagged times. Weather forecasts include temperature, shortwave radiation, and precipitation, provided for different locations within the SES service area (Locarno, Cevio, Piotta, Biasca, and Magadino).

An additional predictor is the historical installed photovoltaic (PV) power, included to capture the progressive impact of distributed PV generation on the observed demand. Since the SES demand data represent the net consumption (i.e., total consumption minus PV self-production), incorporating this feature allows the model to separate the underlying demand evolution from the effect of growing PV capacity.

While for AEM we used as a baseline for comparison the statistical predictor (Derboni et al., 2021) (hereafter referred to as *stat*), for SES we adopted a more standard statistical model based on a seasonal-trend decomposition followed by exponential smoothing, implemented through the STLF function of the forecast package in R (Hyndman and Khandakar, 2008). The method performs an STL decomposition (Seasonal-Trend decomposition using Loess) of the historical demand series, isolating the daily seasonal component. The (deseasonalized) residual component is forecasted using an exponential smoothing state-space model (ETS). The final forecast is obtained by recombining the projected seasonal and ETS-based components. This baseline relies exclusively on past demand observations, modeling only the daily seasonal cycle and excluding any exogenous predictors such as weather variables. Since the STLF approach sometimes fail to correctly capture the seasonal pattern of our time series, likely due to their noisy nature, we also include a naïve method, similar to the statistical approach proposed by Derboni et al. (2021), which computes the average demand at the same hour over the previous seven days.

We also forecast for the water inflow. The goal of the inflow forecaster is to estimate how much water will flow into the reservoir over the next 48 hours, with one prediction for each hour. Since the output size is moderate, due to the lower resolution compared to the demand time series, to do this, we use 48 separate forecasting models (based on LightGBM), each trained to predict the inflow for a specific hour ahead. Instead of predicting the absolute inflow, each model estimates how much the inflow will change compared to the current value.

The models use three types of input data:

- Date and time features, such as the month, and mathematical representations of the day of the year and time of day.
- Past inflow data, including the last 6 hourly measurements before the current time, and summary statistics (average, minimum, maximum) over the past 6 and 12 hours.
- Rainfall forecasts, including predicted rain at the reservoir and average rainfall along the rivers that feed into it. We use both short-term forecasts (1 to 3 hours before each prediction time) and cumulative rainfall over longer periods (2 to 36 hours before each prediction time).

This approach allows us to anticipate how inflow will evolve and support better control of the hydropower system.

For SES, the inflow data are smoothed using a 4-hour moving average to reduce noise before training. The forecasting setup follows the same LightGBM-based framework as for AEM, with an hourly resolution and a 48-hour prediction horizon. The same set of date-time, past inflow, and precipitation features as in AEM is used, with rainfall data obtained for multiple locations covering the basin and the main rivers feeding it. When available, historical precipitation forecasts are used for training; otherwise, historical observations are employed. To reduce overconfidence due to the use of observed rainfall during training (while only forecasts are available at test time), all precipitation-related features are down-weighted by a factor of 0.3.

Forecasting water inflow is a challenging task, and to our knowledge, there are no widely established standard approaches. As a baseline, we compare our predictions with a simple assumption: we consider the inflow to remain constant and equal to the last observed value.

We evaluated the performance of both the forecasting and control strategies applied to the AEM and SES hydropower plants. For AEM, the analysis was conducted over four monthly periods, while for SES, we considered two monthly periods.

The figures show: (A) examples of a forecasts, (B) the average performance across the test periods in terms of RMSE and (C) in terms of sMAPE, all as a function of the forecast horizon.

For AEM, results are shown in Figure 9 for demand and Figure 10 for inflow. For SES, results are shown in Figure 11 for demand and Figure 12 for inflow.

In all cases, the LightGBM-based predictor outperforms the baseline models. This is especially evident for SES. These results confirm that including contextual information such as weather forecasts and installed photovoltaic power is essential to accurately predict both energy consumption and the availability of natural resources, and this can be effectively done by machine learning models. Relying only on periodic patterns and autoregressive features resulted insufficient.

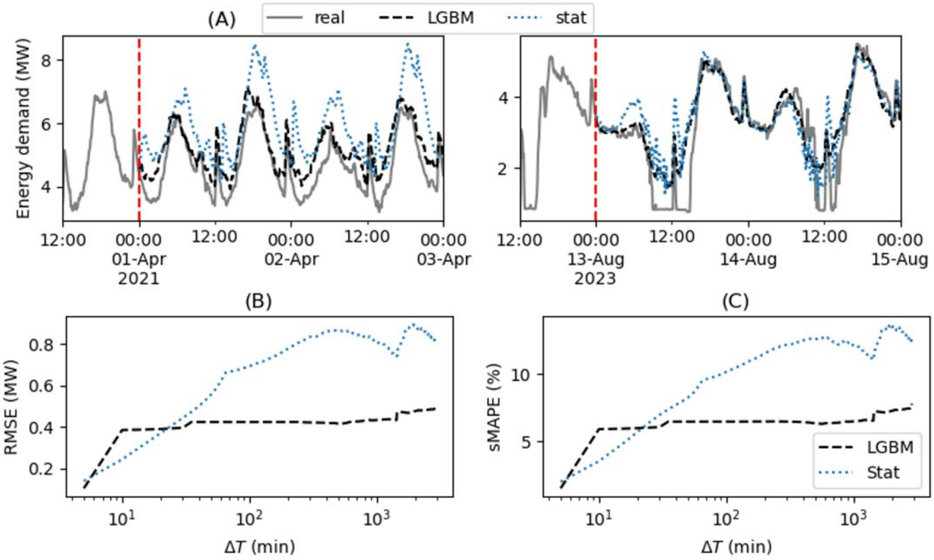


Figure 9: (A) Examples of Energy demand forecasts for AEM. (B) root mean square error (RMSE) and (C) symmetric mean absolute percentage error (sMAPE) of the LGBM and stat predictors.

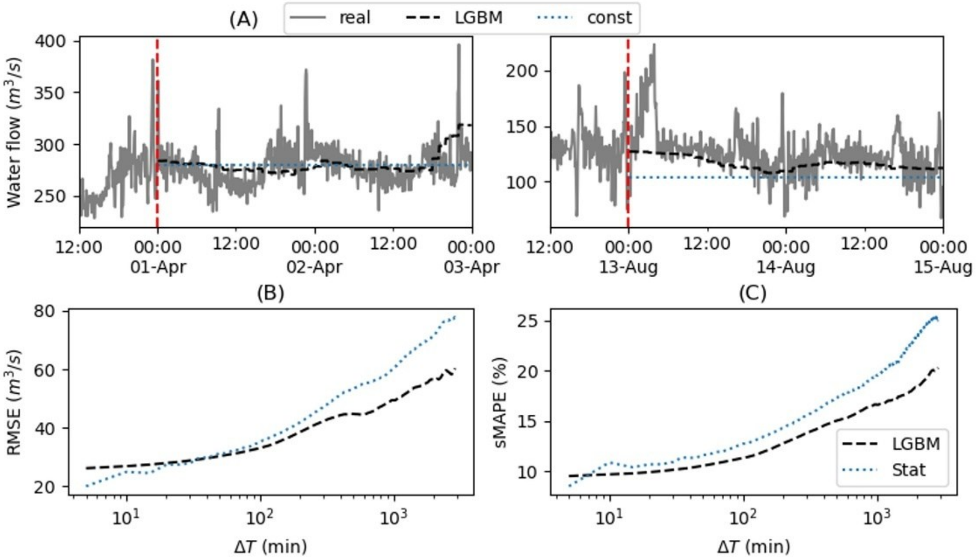


Figure 10: (A) Examples of Water flow forecasts for AEM. (B) RMSE (C) sMAPE of the LGBM and (C) sMAPE of the LGBM and constant flow predictors.

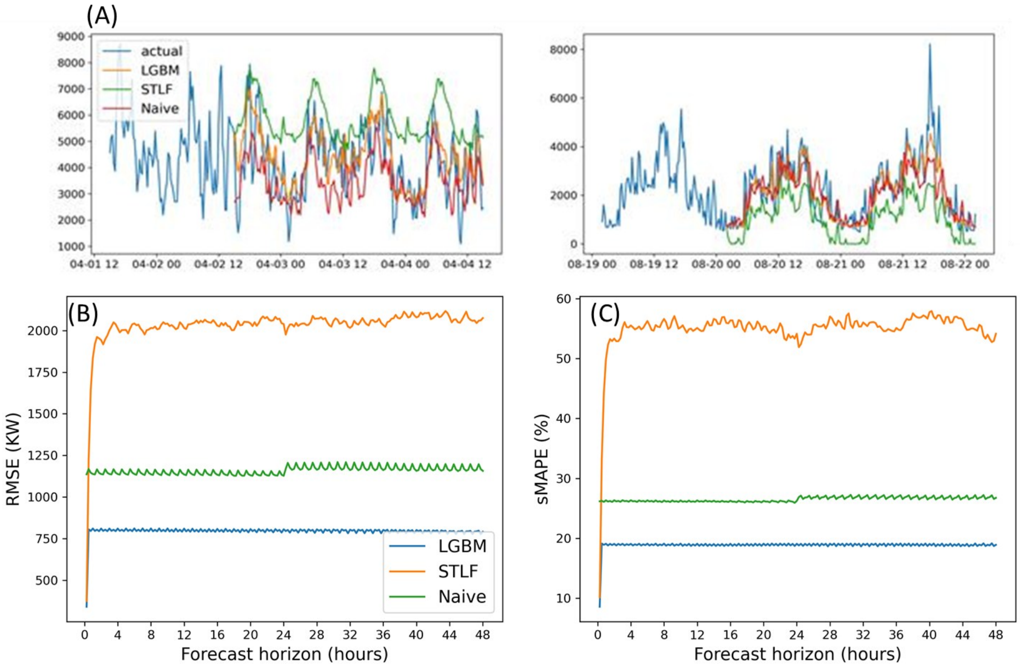


Figure 11: (A) Examples of Energy demand forecasts for SES. (B) root mean square error (RMSE) and (C) symmetric mean absolute percentage error (sMAPE) of the LGBM, STLF and naïve predictors.

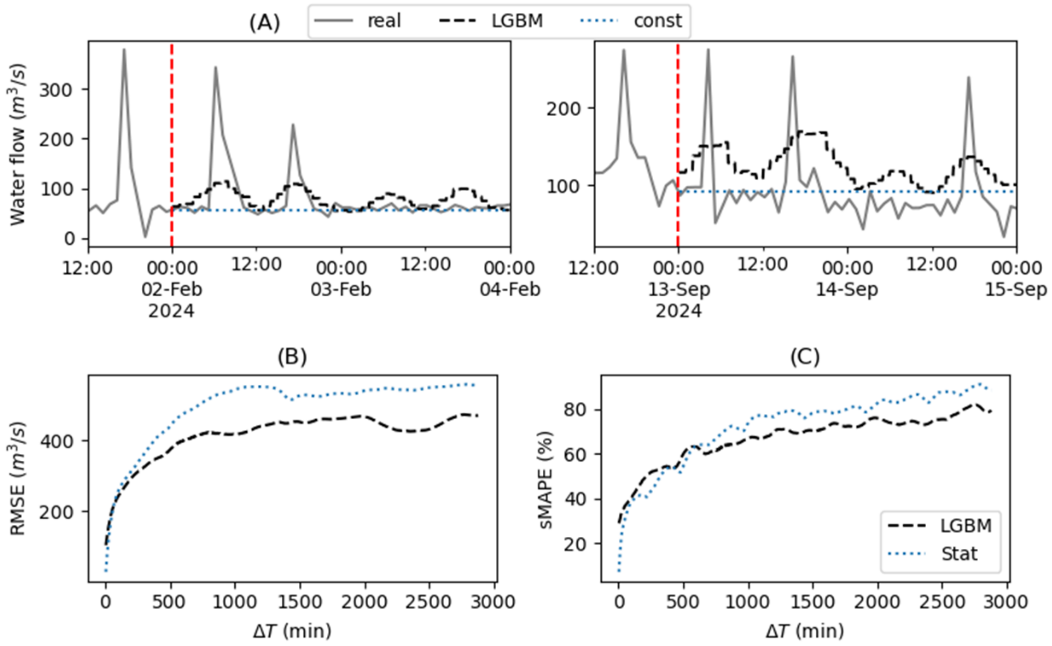


Figure 12: (A) Examples of Water flow forecasts for SES. (B) RMSE (C) sMAPE of the LGBM and (C) sMAPE of the LGBM and constant flow predictors.

The goal of the trained predictors is to provide forecasts upon which the controller can base its decision. In real-time, the predictors receive the most recent time series of demand and inflow, along with weather forecasts and, when relevant, the installed photovoltaic power. Based on this input, they generate forecasts for the next 48 hours. The scheduler then optimizes the hydropower production over this 48-hour horizon, aiming to use the available water primarily during peak demand periods. It returns a production plan for the next 48 hours to the HPP.

In the case of AEM, only the decision for the next 5 minutes is implemented. After that, the optimization is repeated using the updated data from the last 5 minutes, and a new decision is made. SES, on the other hand, follows a different management approach: the scheduling is decided once a day in the morning, for the following 24 hours.

For AEM, we evaluated the quality of the decisions by updating them every 5 minutes. For SES, we tested both this real-time setting (with 15 minute resolution in this case) and a daily scheduling scenario, where the production plan is computed at 10am for the next 24 hours.

Figure 13 shows an example of the dashboard available to the operator. The top panel displays the demand forecast, followed by the inflow forecast (in this case compared with measured values, as it is a simulation). Below, the scheduled production, the resulting net energy demand to be purchased from the cantonal provider, and the water reserves are shown. These last three are compared with the operator’s decisions and their outcomes.

Figure 14 shows the distribution of daily peak demand across the test periods. It compares the operator’s results with those of the scheduler under two conditions: one where it has perfect knowledge of future demand and inflow (referred to as the “oracle”), and one where it uses forecasts from the LightGBM-based models (referred to as the “LGBM”). For AEM we added also the condition where the scheduler receives the demand forecast from the model *stat* and assumes a constant water inflow. We observe that the scheduler consistently reduces the average peak and, in most cases, the maximum peak as well.

Figure 14C also shows that the improvement is clearly reduced when production decisions cannot be updated in real time. This limitation is due to the fact that a dedicated operator is not always available. However, the proposed solution enables optimized automatic control without requiring manual intervention.

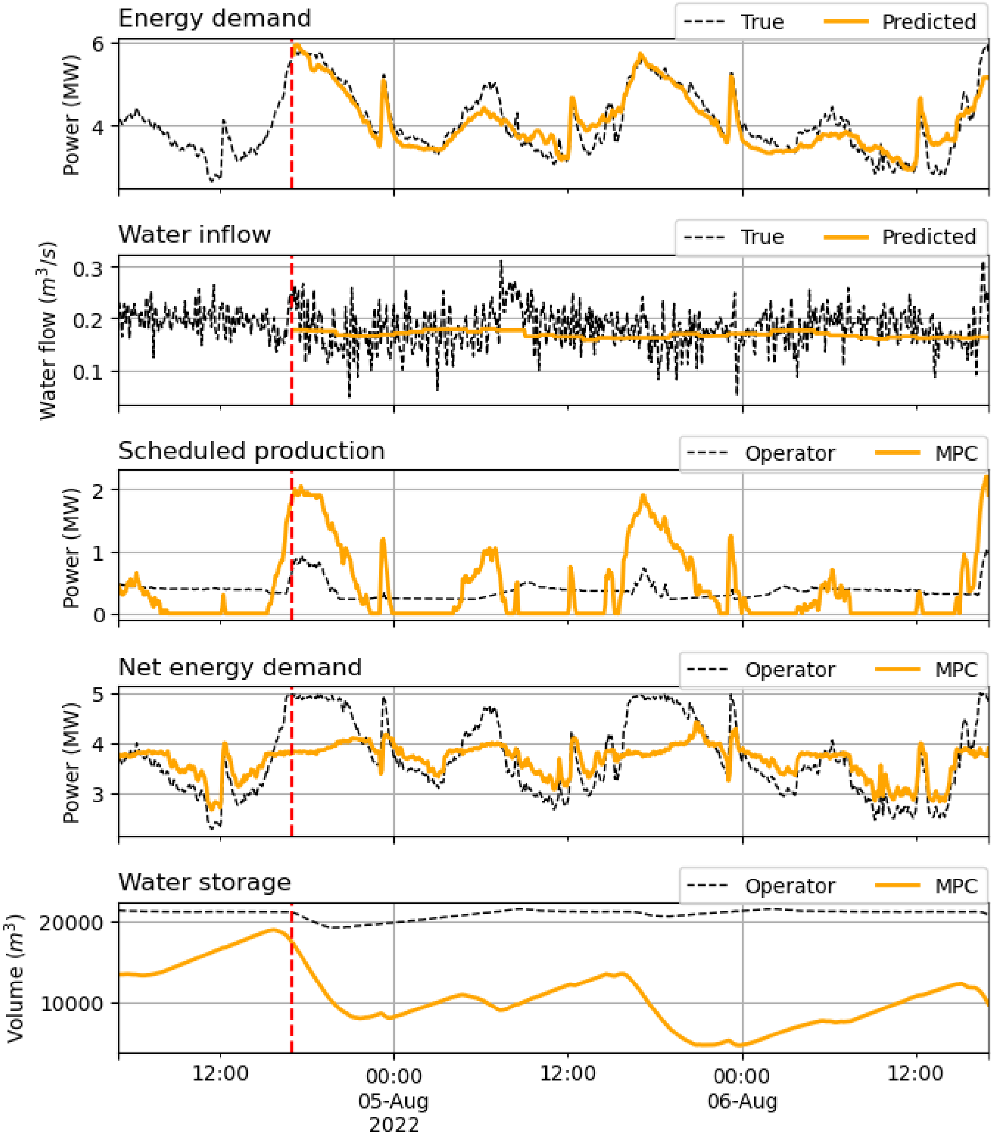


Figure 13: Example of demand and inflow forecasts and scheduling decisions.

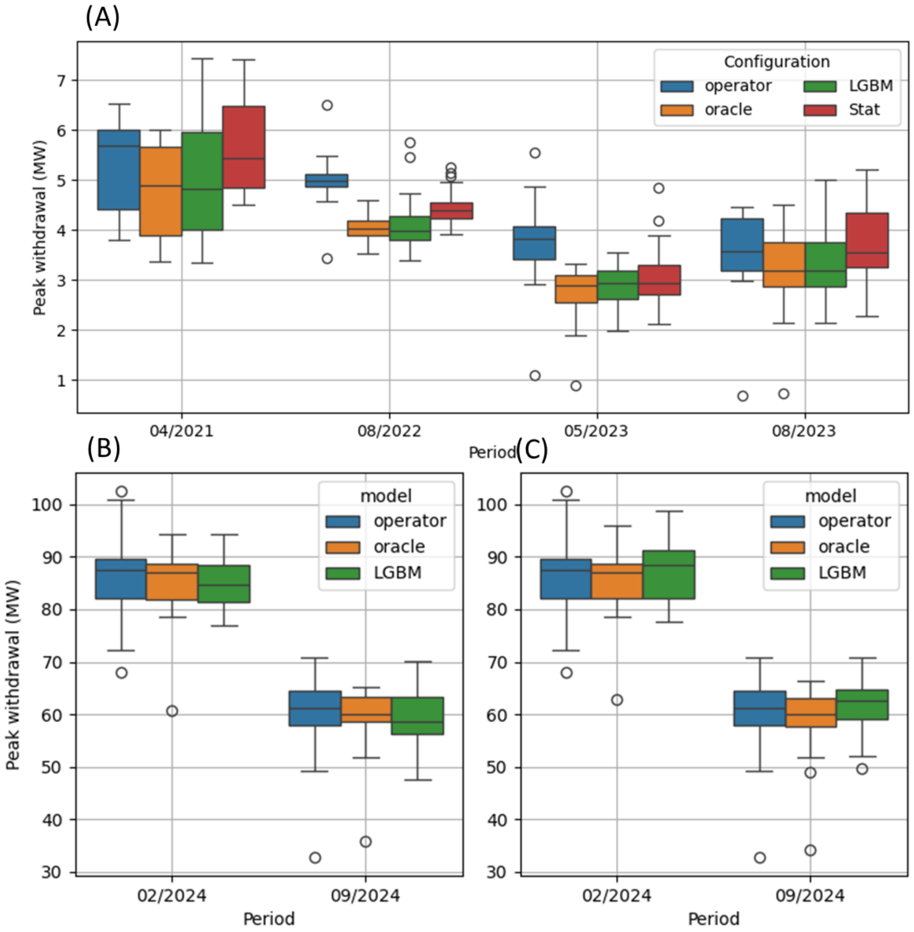


Figure 14: Peak distribution resulting from different scheduling strategies. (A) AEM; (B) SES with decision updated every 15 minutes; (C) SES with decision updated every 24 hours.

According to the information provided on https://www.swissgrid.ch/en/home/customers/topics/tariffs.html#tariffs-and-remunreation—which outlines the tariff structure applied to power consumption and grid services in Switzerland—we assume a power tariff of 50 kCHF per MW. A power tariff refers to the cost charged for the maximum power capacity (or peak load) contracted or used, rather than the energy consumed over time.

Based on this assumption, for the SES case, the scheduler would lead to an estimated saving of about 4 MW (reduction in peak demand) × 50 kCHF=200kCHF during the first period. This saving could increase to 8 MW × 50 kCHF=400kCHF if scheduling is performed every 15 minutes. In the case of AEM, the difference in peak load between the operator and the scheduler is smaller, ranging from −1 MW to +2 MW, with an average of 0.3 MW over the 4 simulated periods, corresponding to an average monthly saving of approximately 16 kCHF.

Battery Sizing in Solar Energy Optimization by MPC. The second application focused on managing a battery system that stores solar energy produced by a small self-consumption community. This was implemented within a simulation environment capable of modeling solar production, self-consumption, and grid demand under varying battery and photovoltaic system sizes.

We perform a set of simulation experiments to support energy communities decision makers to design and evaluate the impact of Photovoltaic (PV) installations and Community battery sizing on self-consumption, i.e. how much of the PV energy produced is consumed locally and self-sufficiency, i.e. how much of the total community demand is covered by local PV + battery under different PV and battery sizing choices and for three seasonal months. Using real consumption data from households in the Lugaggia Innovation Community (LIC), we demonstrated how such simulations—when integrated into the DIGICITIES platform, which provides easy access to historical demand and solar irradiance data—can support optimal sizing of batteries and PV systems according to specific energy management goals.

System and Experiment Setup - We evaluate a community of 10 single-family households aggregated into a single community model. We aggregate the electricity demand which amounts to 19640 kWh, 11672 kWh, and 36764 kWh for the seasonal months of April, August, and December, respectively. The PV installations that are evaluated are 51 kWp (Small), 103 kWp (Medium), 205 kWp (Large) each having a total production of:

PV System Size	April (kWh)	August (kWh)	December (kWh)
Large	24’706	26’941	8’823
Medium	12'353	13’471	4’412
Small	6’176	6’735	2’206

Community stationary battery options:

Battery Option	Energy Capacity (kWh)	Charge/Discharge Power (kW)
No Battery	0	–
Small	20	16
Medium	50	40
Large	100	80
Very Large	200	160

Battery round-trip efficiency: 87.5%.

Demand forecaster - Battery charge and discharge decisions are based on forecasts of community energy demand and photovoltaic production. For the demand forecast, we apply the same LightGBM-based forecasting framework as for AEM and SES, with adjustments to the data resolution and feature set. The demand data have a 5-minute resolution, yielding 576 prediction steps over the 48-hour horizon. Past energy demand features include short-term lags of 5, 10, and 15 minutes, and long-term lags of 1, 2, and 3 days, and 1 week, together with the minimum, maximum, and mean values computed over 10-minute and 30-minute windows around these lagged times. Weather forecasts include temperature, shortwave radiation, and precipitation for the Lugaggia area. Historical weather forecasts are not available, and only the latest forecast (typically issued a few hours before the current time step) is used. Although this may slightly overestimate the available information, it reflects the most realistic data accessible for this case. The baseline model considered are the same as for SES, consisting of an STL decomposition with daily seasonality followed by an ETS model on the residuals and of a naïve approach averaging the demand at the same hour over the last seven days.

Figure 15 shows: (A) examples of forecasts, (B) the average performance across three test periods of one month each, in terms of RMSE, and (C) sMAPE, all as a function of the forecast horizon. Even in this case, using a ML approach that includes covariates (mostly related to weather forecast) provides increased accuracy.

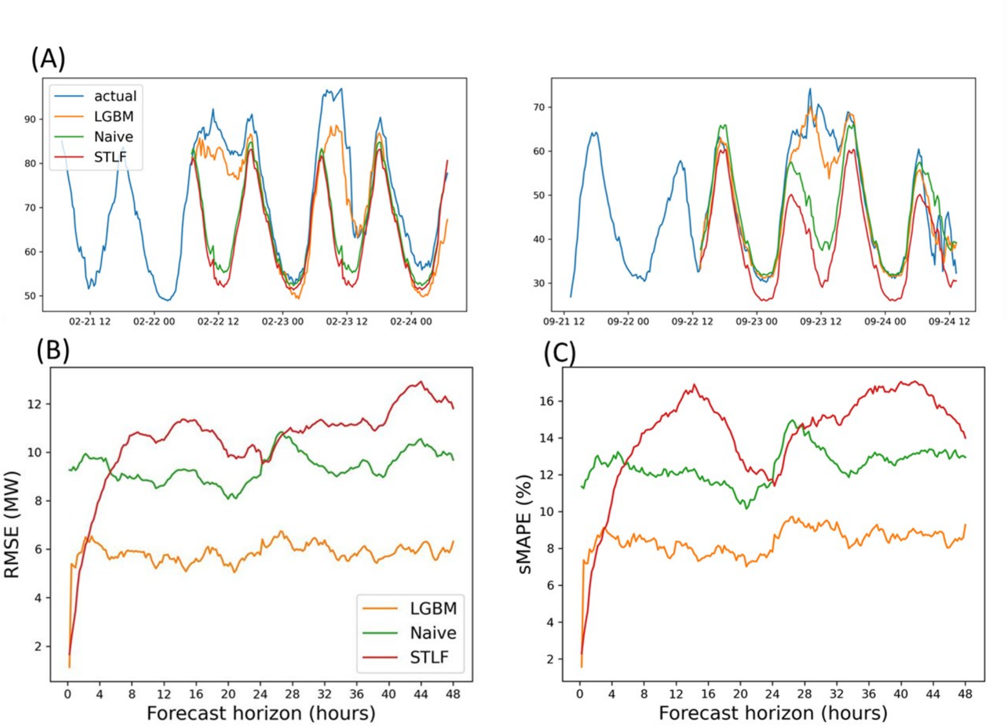


Figure 15: (A) Examples of Energy demand forecasts for LIC. (B) root mean square error (RMSE) and (C) symmetric mean absolute percentage error (sMAPE) of the LGBM, STLF and naïve predictors

For battery sizing, we use energy demand data from LIC households and irradiation measurements collected between 2021 and 2022. Data on PV production was not available and was therefore simulated using the irradiation data. Since weather forecasts for that period were not available from Open-Meteo.com, it was not possible to train an ML model; thus, as a first approximation, PV production was assumed to be known.

Experiments Performed - We tested the following combinations to avoid meaningless combinations: No Battery – Small PV, Medium PV, Large PV to compute the baseline for self-consumption and self-sufficiency

- Small Battery – Small PV
- Medium Battery – Medium PV
- Large Battery – Large PV
- Very Large Battery – Large PV

Every combination is evaluated in the three seasonal months (each totaling 25 testing days) Experiments including a stationary battery are conducted exploiting the MPC controller described in section 3.11. We tested the controller with two demand estimation models LGBM and STLF.
Results – no Battery - We first report the results for the baseline case, without a battery system. In this case, as there is no battery, no control decision can be taken, therefore no demand prediction is necessary. The following table reports the Self-consumption (sc) and self-sufficiency (ss) percentages.

Table 3: Self-consumption (sc) and self-sufficiency (ss) for the no battery scenario.

PV System Size	April (sc/ss)	August (sc/ss)	December (sc/ss)
Large	30%/38%	26%/59%	67%/16%
Medium	50%/32%	46%/53%	92%/11%
Small	78%/24%	76%/44%	100%/6%

We immediately observe that, for the given community, the sole installation of a shared PV community installation would greatly impact both self-consumption and self-sufficiency KPIs.
Of course, the impact is not homogeneous across seasonal months, whit April better balanced between self-consumption and self-sufficiency. In August, when the energy demand is much lower, self-sufficiency increases. Instead, in December, when PV production is low and demand is quite high, a large portion of the production is locally consumed.

Results –Small Battery, Small PV

Table 4: Self-consumption (sc) and self-sufficiency (ss) for the Small - Small scenario

Demand forecast	April (sc/ss)	August (sc/ss)	December (sc/ss)
LGBM	84%/26%	83%/47%	100%/6%
STLF	82%/26%	82%/47%	100%/6%

We observe that the introduction of a small-sized community battery increases self-consumption with the capability of storing excess production to some extent. The more accurate LGBM model allows to capture more energy to be used when necessary.
Similarly, self-sufficiency is slightly improved thanks to the battery. We observe that in December no improvement is obtained as the entire production is consumed locally even without a stationary battery.
For this scenario, we report the total energy exchanged (e) with the stationary battery and the estimated number of cycles (c).

Table 5: Total energy exchanged (e) with the stationary battery and estimated number of cycles (c) for the Small - Small scenario.

Demand forecast	April (e/c)	August (e/c)	December (e/c)
LGBM	414 kWh/21	528 kWh /26	0 kWh /0
STLF	403 kWh /20	504 kWh /25	27 kWh /1

We observe that the community battery is used in April and August with an average of one cycle per day. Some control errors can be spot in December when using the STLF model as some energy is stored in the battery but that energy did not come from the PV installation. In this case the model predicted an excess of production (lower demand) leading the controller to store some energy.

Results – Medium Battery, Medium PV

Table 6: Self consumption (sc) and self-sufficiency (ss) for the Medium - Medium scenario

Demand forecast	April (sc/ss)	August (sc/ss)	December (sc/ss)
LGBM	59%/36%	55%/62%	95%/11%
STLF	59%/36%	55%/61%	96%/12%

With medium battery and medium PV we observe a more marked increase in self-sufficiency. That indicates that the system is capable of storing excess energy and use it when necessary, during periods with higher demand.
For this scenario, we report the total energy exchanged (e) with the stationary battery and the estimated number of cycles (c).

Table 7: Total energy exchanged (e) with the stationary battery and estimated number of cycles (c) for the Medium - Medium scenario.

Demand forecast	April (e/c)	August (e/c)	December (e/c)
LGBM	1173 kWh/23	1292 kWh /26	104 kWh /2
STLF	1204 kWh /24	1250 kWh /25	289 kWh /6

For what concerns, exchanged energy we have similar observations. Battery is used daily in April and August up to its capacity. STLF model performs forecast mistakes in December leading the controller to use much more energy than really produced in excess.

Results – Large Battery, Large PV –

Table 8: Self consumption (sc) and self-sufficiency (ss) for the Large - Large scenario

Demand forecast	April (sc/ss)	August (sc/ss)	December (sc/ss)
LGBM	40%/48%	35%/78%	83%/19%
STLF	40%/48%	35%/75%	82%/19%

In the scenario with Large PV and large Battery, we observe that the PV production is massive in April and August and the self-consumption metric is expected to be lower. The excess of production cannot be stored even if the battery is of a decent size (an average of 10kWh per household).
We also observe a neat increase in self-sufficiency as the available energy is now better stored and distributed when needed.

For this scenario, we report the total energy exchanged (e) with the stationary battery and the estimated number of cycles (c).

Table 9: Total energy exchanged (e) with the stationary battery and estimated number of cycles (c) for the Large - Large scenario.

Demand forecast	April (e/c)	August (e/c)	December (e/c)
LGBM	2506 kWh/25	2706 kWh /27	1469 kWh /15
STLF	2539 kWh /25	2510 kWh /25	1547 kWh /15

For what concerns, exchanged energy we now observe that there is exploited excess production even in winter season. For April and August we see that the battery capacity is fully exploited with the number of cycles platooning at about 25.

Results – Very Large Battery, Large PV

Table 10: Self consumption (sc) and self-sufficiency (ss) for the Very Large - Large scenario

Demand forecast	April (sc/ss)	August (sc/ss)	December (sc/ss)
LGBM	49%/57%	42%/91%	92%/21%
STLF	48%/56%	41%/84%	92%/22%

In the scenario with Large PV and Very Large Battery, we observe that we can further benefit from the large PV production as now it can be stored in a larger battery and exploited to improve self-sufficiency. In august almost the entire energy needs are covered by renewable energy.
For this scenario, we report the total energy exchanged (e) with the stationary battery and the estimated number of cycles (c).

Table 11: Total energy exchanged (e) with the stationary battery and estimated number of cycles (c) for the Very Large - Large scenario.

Demand forecast	April (e/c)	August (e/c)	December (e/c)
LGBM	4777 kWh/24	4523 kWh /23	2237 kWh /11
STLF	4818 kWh /24	4317 kWh /22	2484 kWh /11

We now observe that the large capacity of the battery is not fully exploited in August because the consumption is not sufficient to empty it the next day.

Conclusions - This simulation study for a 10-household energy community demonstrates a clear and significant relationship between PV sizing, battery capacity, and seasonal energy patterns. The simulation tools developed within DIGICITIES allow for an informed choice on Stationary Battery and Photovoltaic installation in an energy community. The results provide critical insights for decision-makers aiming to balance self-consumption (sc) and self-sufficiency (ss).
The baseline "no battery" scenarios confirmed the fundamental challenge: without storage, the community either fails to consume its own production during high-generation months (like August, with low sc) or fails to meet its demand during low-generation months (like December, with low ss).
The introduction of a community battery mitigates these seasonal imbalances by enabling energy time-shifting. Small Batteries (with Small PV) primarily increase self-consumption by capturing immediate excess energy that would otherwise be lost. Medium and Large Batteries provide a much more significant boost to self-sufficiency, allowing the community to store and utilize larger amounts of energy during periods of high demand. Diminishing returns were observed in the "Very Large Battery" scenario. While it achieved the highest peak self-sufficiency (91% in August), the battery's capacity was not fully exploited, suggesting a potential for over-sizing and unnecessary capital expenditure. The "Large PV / Large Battery" combination appeared to offer a more balanced solution, as the battery capacity was consistently and fully utilized. Finally, the accuracy of the demand-forecasting model was shown to be important for operational efficiency. Although the more accurate LGBM model only matched or slightly improved performance, it, more importantly, avoided the control errors observed with the STLF model, which occasionally led to improper battery charging from the grid in winter.

These experiments show that a shared battery is a critical component for maximizing the value of a community PV installation. However, decision-makers must carefully size the battery relative to both the PV capacity and the community's seasonal demand profile to find the most cost-effective balance between self-consumption and self-sufficiency. This, along with optimal battery management, requires the availability of large retrospective datasets to train accurate predictors and run realistic simulations.

Integration of Data and Models in DIGICITIES. The development of this use case supported the design of DIGICITIES ontology to accommodate the data and models required for advanced energy resource scheduling applications and their exchange with an external analyst and service provider.

Figure 16 illustrate the flow of information for the training of the predictive ML models and the real time operation of the scheduler.

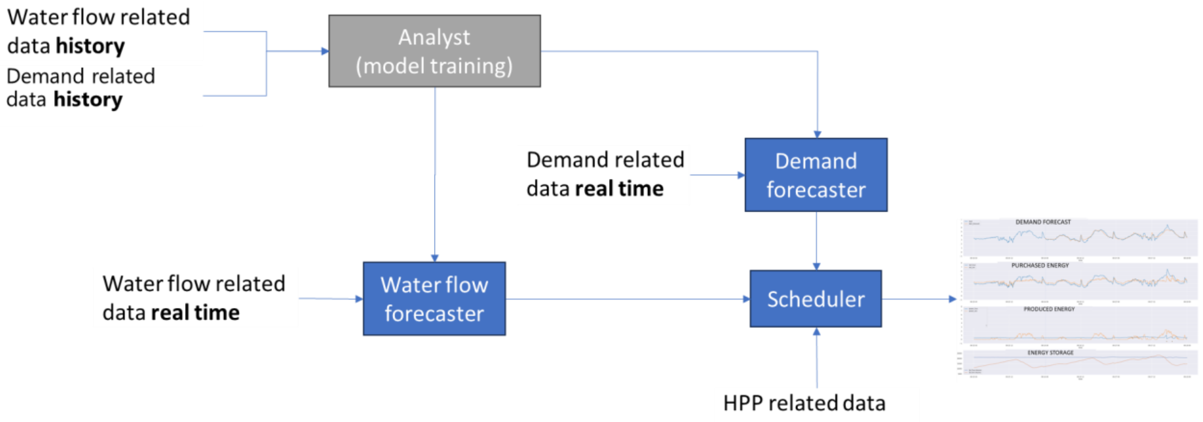


Figure 16: information flow within the HPP scheduling system.

A structured process was designed in which data managed by DIGICITIES for all the applications of this use case are organized around five main entities:

- Renewable energy source (e.g., solar production or hydro reservoir)
- Electricity grid
- Controllable asset (e.g., battery or hydroelectric power plant)
- Weather (grouping variables relevant to forecasting demand and renewable availability)
- Machine learning model (providing metadata and access to trained models used for control)

Each entity is described through parameters and data sources. For example, the hydro reservoir entity includes volume limits and historical inflow data, while the hydro turbine includes maximum rated power and historical production. These parameters are used by the scheduling service to configure operational constraints and locate relevant data sources—either historical for training or live for real-time operation.

The JSON structure used to call the scheduling service includes references to these entities, specifying both static parameters (e.g., reservoir capacity, turbine limits) and dynamic data sources (e.g., FTP endpoints for live inflow or demand). During the training phase, analysts use historical data to develop predictive models, which are then registered and organized within DIGICITIES. During operational use, the service call includes both the live data sources and the reference to the trained model appropriate for the scenario.

This approach enables flexible, ontology-driven integration of data and models, supporting scalable deployment of predictive control strategies across different energy systems.

Table 12: Summary of Impact Assessment and KPIs for Swiss UC1

Impact Category	KPI	Result
Data Access	Number and size of time series used for training the models	AEM - from Dec 2018 to Sept 2023: 8 time series. SES – from Jan 2018 untill now: 40 time series (mostly weather forecasts) LIC - from 2021 to 2022: 4 time series
	Number and quality of new data available (forecasts)	3 demand forecasts, 2 inflow forecasts. Prediction errors are in Figures
	Time for training data preparation and validation set up	About 6 FTE months for the first demand and inflow models (AEM). About 2 months for SES models and two weeks for LIC model.
Decision making	Comparison of historical data and predictions	Figures from 30 to 34
	Comparison of operator and controller decisions	Figure 35 and 36
	Support provided in the choice of the battery size	Quantification of self-consumption and self-sufficiency for different battery and PV sizes - Tables 5 to 13
Energy Impacts	Peak reduction	Figure 36
	Energy cost reduction due to the reduced size of monthly peaks	AEM: average monthly reduction of 16kCHF (estimated over 4 moths) SES: average monthly reduction between 96 and 220kCHF (estimated over two months)
	Increased self-consumption of the self-consumption community exploiting optimized battery management	From Table 3 to Table 11

6.1.2 Use Case 2 – Urban Energy Planning

The energy planning need-owners, a utility company and a municipal planner, provided reports and data from past energy modelling studies. In addition to this, they also shared the configuration files for their inhouse energy model. The goal was to accommodate these processes using the DIGICITIES platform.

The data and the reports helped the development of the ontology to accommodate their use case. In addition to this, the configuration files help us develop ways to describe data requirements so that the modelling needs could be satisfied through DIGICITIES. The data was processed and transformed to be suitable for storage on the NextCloud. This involved storing timeseries into individual files. A knowledge graph of the existing data was created using the Excel template. Sympheny programmed an API for a regional energy planning, and this was directly integrated into the API module of the platform. A service requirements YAML template was generated that referenced the components and attributes in the ontology, the inputs are shown in Figure 17.

```
service_name: Sympheny-Regional-Analysis
scenario_data:
  uri: Scenario.URI
  name: Scenario.label
  Region:
    name: Region.label
    uri: Region.URI
    EnergyConsumer:
      link: CL.Region.EnergyConsumer
      template:
        uri: EnergyConsumer.URI
        ElectricityDemandProfile: EnergyConsumer.ElectricityDemandProfile.hasHistoricTimeSeriesReference
    EnergyConverter:
      link: CL.Region.EnergyConverter
      template:
        uri: EnergyConverter.URI
        CAPEX: EnergyConverter.CAPEXPerRatedPower
        Efficiency: EnergyConverter.Efficiency
    EnergyGenerator:
      link: CL.Region.EnergyGenerator
      template:
        uri: EnergyGenerator.URI
        SolarPotentialProfile: EnergyGenerator.Power.hasHistoricTimeSeriesReference
    EnergyCarrier:
      link: CL.Region.EnergyCarrier
      template:
        uri: EnergyCarrier.URI
        EnergyCost: EnergyCarrier.EnergyCost
```

Figure 17: Service YAML for the Sympheny Regional Energy Planning

The need-owners also expressed a need to obtain geospatially resolved demand data for EV and solar PV facades that could be easily input into the service. This formed a set of distinct sub-services that could then feed the main Regional Energy Planning service. The relationship of these services is shown in Figure 18.

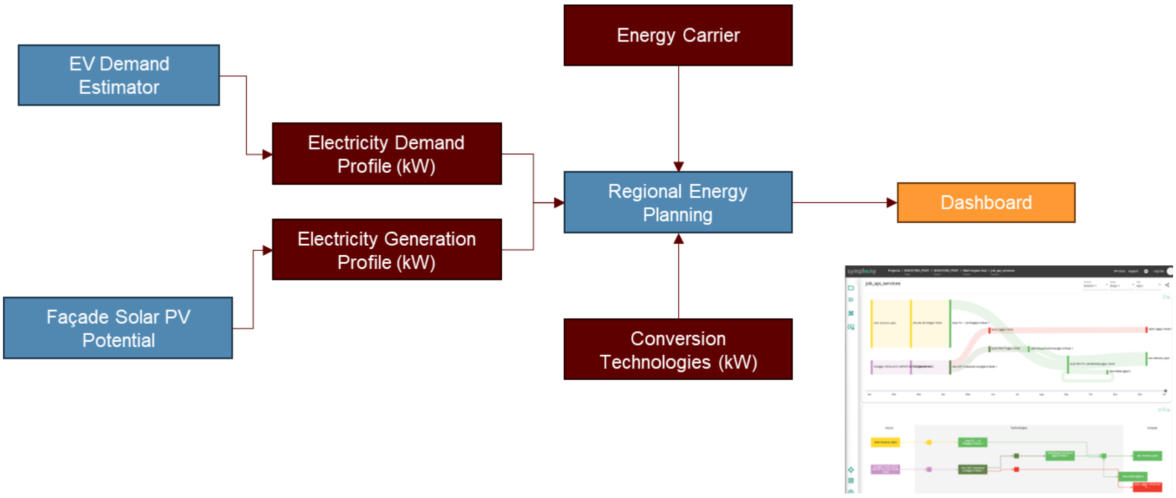


Figure 18: An overview of the inputs and services enabled in the DIGICITIES platform for the Energy Planning Usecase

Table 13: Summary of Impact Assessment and KPIs for Swiss UC2

Impact Category	KPI	Result
Data Access	Quantity of data sources used in the energy planning study (comparison with existing study)	We were able to provide new data sources for façade and EV potential.
	Time saved accessing and processing data (estimate)	Not possible to evaluate.
	Time saved creating usage agreements between partners (estimate)	The middleware with TrustRelay was configured but not integrated into the final platform. This is work for future relevance.
Decision making	Accuracy of projected baseline (comparison with historic data)	Not possible to evaluate.
	User evaluation	User evaluation of the platform was determined by survey. See Section 6.2.
	Stakeholder reach	Only project members
	Metrics reported in the energy planning report	Not possible to evaluate.
Energy Impacts	Percentage of self-consumption achieved	Not possible to evaluate.
	Carbon intensity of future scenarios	Not possible to evaluate.

The impact assessment for the energy planning use case, Table 13, may suggest limited success of the project in terms of impact; however, several important factors should be considered before making this assessment. Many of the KPIs listed in Table 2 require quantitative before-and-after comparison, but a meaningful baseline could not be identified against which to compare. Current municipal energy planning processes do not routinely validate projections against historical data or produce standardised metric sets that would allow direct comparison with DIGICITIES outputs. Attempting to quantify these KPIs under such conditions would have been misleading rather than informative. Similarly, the energy impact KPIs relating to self-consumption and carbon intensity require operational deployment over a sufficient time horizon; the project scope covered platform development, integration, and piloting rather than long-term operational monitoring, and these remain relevant targets as the platform moves beyond the pilot phase. Where evaluation was possible, it was carried out: the platform successfully demonstrated the creation of new data sources for façade and EV potential not available in the existing study, and user evaluation was conducted via stakeholder survey.

The project's priority was developing a platform that could satisfactorily accommodate all use cases, and this is reflected in the stakeholder responses. Direct users of the platform, such as energy planners, were skeptical of the benefits and would need concrete use cases demonstrating value to their operations before adoption. Barriers to implementation were also a concern, meaning it needs to be clear how these could be overcome before widespread implementation is possible. Digital service providers saw more potential for their businesses, as they are closer to the challenges DIGICITIES is designed to accommodate, such as scalable data processing and interoperability improvements. This was also recognized by the service provider specializing in data marketplaces and data products. Current value chains feature misaligned incentives regarding data sharing, with operators viewing operational data as commercially sensitive whilst service providers may wish to maximise financial compensation for providing data. To achieve its full potential, regulatory mandates for standardised formats, financial incentives for semantic data creation and publication, and liability frameworks that address data sharing risks are all required.

6.1.3 Other Usecases

The process in DIGICITIES is highly adaptable and there are several other use cases are envisioned to be supported by the platform. This includes the energy dashboard developed by the Austrian partners and the Digital Twin Platform of the Horizon REFORMERS project. In addition to this, the platform could also be integrated into the NESTCloud environment of NEST demonstrator at Empa, and also serve the energy modelling community of SWEET Cosi. These activities have already helped develop the ontology to make it applicable to a diverse range of applications and could provide a solid basis for developing and maintaining the platform and ontology beyond this project.

6.2 Stakeholder Assessment

The stakeholder assessment considers the perspective of the two main stakeholder groups of the platform, users of the interface and digital service providers.

6.2.1 Need Owners

Results in HPP peak shaving showed that automatic scheduling of the power production is feasible and allows better energy management which for the utilities results in a reduced energy price.

We have set up a testing phase of the scheduler at SES. Updated time series of energy demand and reservoir level will were sent to us by SES daily. These data were fed into the pre-processing pipeline and the forecasting models, which generated 48-hour predictions for both demand and inflow. These forecasts were then be provided to both the scheduler and SES, along with the scheduler’s production decisions. The results during this phase were aligned with the forecast accuracy and peak shaving performance internally assessed.

The energy planning use case stakeholders were surveyed after a demonstration meeting. This is an analysis of the findings:

- **Data infrastructure improvements were more appealing than modeling efficiency**. Interest focused on standardised terms, ingestion, traceability, and workspace management.
- **Migration challenges and costs were a concern and the primary adoption barrier.** Time to transition existing workflows currently outweighs anticipated benefits.
- **Concrete proof needed before commitment.** Case studies and demonstrated time savings are essential to drive adoption decisions.

6.2.2 Digital Service Providers

The main digital service providers saw opportunity in working with DIGICITIES. A summary of their experience is provided below:

- **Reduces data integration workload.** Potential to streamline the volume of data required for platform integration.
- **Accelerates user data preparation.** High potential to reduce time users spend on data setup and formatting.
- **Expands market reach through simplification.** Simplified services could unlock access to new client segments e.g. offering consistent reporting.
- **Increases confidence in API data quality.** Standardized inputs provide greater assurance in data sent to services.
- **Improves API efficiency at scale.** Enables one API to handle multiple scenarios more effectively.

7 Discussion

This section discussed the overall project approach, explores the innovative concepts, answers the research questions in the proposal and gives a perspective on future research and activities that could build on the results of this project.

7.1 Innovative Platform Concepts

7.1.1 Extensible Ontology

The DIGICITIES ontology framework offers substantial advantages for application developers compared to traditional domain ontology adoption. Conventional energy ontologies such as CIM, CGMES, or SAREF present significant barriers: developers must invest considerable time understanding complex taxonomies, navigate overlapping terminology across multiple standards, and encounter situations where specific use case requirements do not align with predefined structures. Maintenance burdens compound these challenges as external ontology updates can break existing implementations, whilst importing multiple subontologies creates unmanageable conflicts and versioning issues.

The DIGICITIES approach fundamentally inverts this paradigm by prioritising service-specific data requirements. Developers specify only the components and attributes their services require, creating lightweight, purpose-built extensions rather than adopting comprehensive domain models containing irrelevant concepts. The attribute-centric design provides consistent structural patterns for diverse data types physical measurements, costs, performance curves, temporal events whilst maintaining semantic precision through QUDT and FIBO integrations.

Critically, internal flexibility does not preclude external interoperability. The systematic mapping framework enables developers to establish equivalence, hierarchical, or similarity relationships with established standards when integration becomes necessary, bridging internal models with external ecosystems on-demand rather than imposing universal conformance. This proves particularly valuable in multi-standard environments requiring simultaneous SAREF and CIM compliance.
The accompanying DIGICITIES platform amplifies these benefits by abstracting RDF/OWL complexities behind intuitive interfaces. The Ontology Manager enables extensions without mastering semantic web technologies, the Replica Builder populates knowledge graphs through Excel imports, and the Service Registration module defines YAML-based requirements with automatic validation. The Scenario Builder enables rapid analytical variant configuration, whilst the API Submission module handles semantic-to-JSON/YAML conversion, eliminating custom transformation logic. This integrated tool chain transforms semantic modelling from specialised expert activity into accessible workflow, enabling developers to leverage knowledge graph capabilities whilst focusing on domain logic, ultimately reducing time-to-deployment and maintenance costs.

7.1.2 Semantic Data Products

Semantic data products represent a structured approach to data publishing that addresses fundamental challenges in energy data sharing: inconsistent formatting, ambiguous semantics, and disconnected metadata. Each DIGICITIES data product comprises two integrated components: a semantic graph containing instance data conforming to the ontology structure alongside any necessary ontology extensions, and mandatory resource files stored in cloud infrastructure including geospatial data (GeoJSON), time series datasets (CSV), or domain-specific standardised formats such as EPW weather files. The semantic graph provides formal definitions and relationships between components whilst resource files contain the operational data itself. Critically, resource files are referenced within the graph through explicit links, establishing provable connections between semantic descriptions and actual data assets. The components of a DIGICITIES data product are shown in Figure 19.

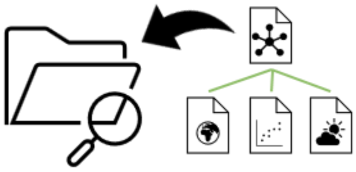


Figure 19: Components of a semantic data product

This architecture offers advantages over conventional data distribution approaches. Publishers package data with machine-readable semantics, eliminating ambiguity about units, temporal characteristics, spatial references, and attribute meanings that cause problems in ad-hoc CSV or spreadsheet exchanges. Consumers can programmatically validate data product completeness and semantic compatibility before integration. The Scenario Builder leverages this structure to enable direct ingestion of data product instances into scenario configurations, with automatic validation against service requirements ensuring only compatible data products, comprised of the correct components and attributes, can be selected. This reduces data preparation overhead compared to manual schema mapping and transformation workflows.

The separation between semantic descriptions and resource files supports flexible deployment patterns. Graphs can be queried independently for discovery and validation without downloading large time series or geospatial datasets, whilst resource files can be cached, versioned, or stored in optimised formats. A download client component enables third-party services to programmatically access data products ingested into workspaces, retrieving both semantic metadata and associated resource files through authenticated requests. This facilitates automated data pipelines where external analytical tools consume DIGICITIES data products without manual file transfers.
For data publishers, this structure provides a standardised publication framework ensuring datasets are discoverable, interpretable, and useable. For consumers, semantic data products eliminate data wrangling bottlenecks, enabling focus on analysis rather than data preparation.

7.1.3 Decentralised Data Sharing

The integration of Keycloak federated identity management within the DIGICITIES platform architecture presents significant opportunities for future development toward truly decentralised data exchange paradigms. Whilst the current implementation employs centralised NextCloud storage infrastructure for all workspace artefacts, a federated storage model could enable users to connect their own cloud storage services as direct replacements for NextCloud. This architectural evolution would allow users to authenticate through their organisational identity providers via Keycloak, which would then manage credentials for their connected storage backends—whether Azure Blob Storage, AWS S3, Google Cloud Storage, or private infrastructure. The entire workspace, including knowledge graph replicas, ontology extensions, scenario configurations, data products, and historical service results, would reside exclusively within the user's connected storage environment rather than platform-controlled infrastructure.

This approach would fundamentally reconceptualise the platform's data governance model, transitioning from centralised data custody to distributed architecture where users maintain absolute sovereignty over their digital twin assets. Platform services would operate as computational orchestration layers, executing TTL parsing, scenario generation, and API submissions against user-controlled storage endpoints through Keycloak-managed authentication credentials. Platform administrators would possess no direct access to user data, facilitating semantic interoperability and analytical workflows without data custody responsibilities. All read and write operations performed by platform modules would target user-specified storage locations rather than shared platform repositories.

This architectural pattern aligns closely with dataspace principles articulated by the International Data Spaces Association and European Commission frameworks, which emphasise data sovereignty and secure cross-organisational exchange through standardised protocols whilst maintaining owner control. Users could selectively share specific data products or scenario configurations with collaborators through native cloud storage sharing mechanisms, implementing granular access policies at the storage layer. GraphDB integration could similarly evolve toward federated architectures, where triple stores are deployed within user-controlled infrastructure and accessed through authenticated connections rather than centralised repositories.

7.2 Answering the Original Research Questions

This section revisits the original research questions and tries to answer as much as possible based on the outcome of the project.

7.2.1 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 under-utilised data resources are available to utilities and municipalities to make more accurate demand projections and better decision-making?

  We found that utilities and municipalities are using a diverse amount of data for energy planning; however, they often need to transform it for their models. We found that a large amount of time is dedicated to preparing demand calculations and it is not possible to validate or check whether these are accurate.

- **RQ T.2** What are the key semantic attributes of the data needed to facilitate data exchange across the digital value chain?

  We have focused on the defining attribute categories to make it easier for services to define their requirements. These attributes can be linked to different components giving full flexibility in defining data requirements in the digital value chain.

- **RQ T.3** How can machine-learning techniques be used to address challenges of data incompleteness and inaccuracy?

  ML approaches to improve data quality has not been incorporated as a standard feature on the DIGICITIES platform. The reason for this is the focus on semantic interoperability. ML applications could be used in the production of data products for integration onto the platform. But it is out of scope.

  However, the LightGBM models used in this project can handle missing data natively (as long as the gaps are not too long), making them suitable for incomplete datasets without necessarily resorting to imputation. This approach is efficient and avoids the arbitrary choice of an imputation model. Moreover, the trained models can, if needed, be used to impute missing values themselves. We conducted some experiments and found that, in the context of time series, this strategy is both more computationally efficient and more accurate than extensions of traditional tabular data imputation methods, such as multiple imputation implemented in the Amelia package.

- **RQ T.4** How can federated queries be applied to multiple heterogeneous urban data sources and used to project the energy demand of buildings?

  The concept of federated queries was originally intended to query endpoints of external data for integration into the platform. Such federation of queries was demonstrated through interfacing with the TrustRelay middleware. Internally, the platform used named graphs to separate data for independent querying.

- **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?

  We determined that the key aspects to consider in such architectures is decentralization and the ability to connect individual storage accounts to the platform. This should be served by federated identity management.

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

  The DIGICITIES platform provides data structures for future projections that could be used to inform decision making. The processes established in the hydro scheduler demonstrate how DIGICITIES could inform such processes.

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

  The creation of a Digital Twin was out of scope of this project due to the availability of data. We now have the tools to accommodate real-time data and feed services that could synchronize physical assests as part of a Digital Twin, however this requires more work on.

7.2.2 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?

  There are initiatives in Swtizerland, such as the Swiss Data Alliance, that clearly demonstrate the value adoption of standardised models for energy exchange in domains such as urban energy. However, ontologies and data structures are only part of the solution. There needs to be a data economy with tools supporting governance and access. We have shown the role DIGICITIES could play in such a data economy.

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

  There formalisation of scenarios for the energy perspective and strategies for the Energy Transformation would be useful for consistent evaluation of scenarios across municipalities. DIGICITIES could provide the tools to achieve this.

7.2.3 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?

  Energy planners prioritize data management capabilities - standardized terms, traceability, and improved data confidence. Service developers seek reduced integration workloads, time savings, and API efficiency for scalable scenarios. Both require validated case studies and quantified time savings before commitment, with technology providers additionally motivated by market expansion opportunities.

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

  Essential features include documented case studies from similar organizations, quantified time savings metrics, and structured training. Plans must address migration concerns, workflow transitions, and tool compatibility. Concrete demonstrable cases are critical to overcome adoption challenges and scepticism.

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

  The platform offers moderate rather than transformative improvements, with migration costs outweighing perceived benefits for some stakeholders. Technical benefits include enhanced data quality and API efficiency, but compatibility concerns and resistance to process changes create implementation challenges. Impact varies significantly by stakeholder type and existing infrastructure.

- **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?

  Citizens were not considered in this project, so it is not possible to answer.

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

  The architecture was perceived to require significant workflow adaptation, with migration burden as the primary barrier. Manual configuration and data integration practices could see moderate improvements, though transformative impact remains uncertain. Standardization and traceability represent departures from contractor-dependent methods, while developers indicate substantial efficiency potential in data preparation and API management.

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

  Researchers were not considered in the stakeholder engagement, but researchers could greatly benefit from formalised representation and the traceability of a scenario's assumptions. There are also several research projects, such as Sweet Cosi, that could benefit from the data modelling approach in DIGICITIES. Management of ontologies and extensive semantic reasoning are considered key areas for future research.

8 Conclusion and outlook

The DIGICITIES platform has demonstrated how need-owners in the digital value chain could be connected using linked semantic layers. The integrated platform, with modules covering ontology management, digital replica building, scenario construction, and API submission, provides a toolbox for generating consistent graphs representing energy systems. Data exchange is envisioned using semantic data products comprised of a mandatory graph structure and optional reference files containing standards formats such as timeseries and geospatial data. It also allows for the referencing of specialized standard formats such as weather files, commonly used in physics-based simulation, which provides flexibility for diverse service requirements. However, translating this technical proof-of-concept into widespread adoption requires proving the value of the platform in complex energy planning case studies to overcome foreseen adoption barriers.

While both use cases were evaluated in terms of platform functionality and stakeholder feedback, quantitative impact assessment was only fully achievable for the hydroelectric scheduling use case, where a suitable operational baseline existed. For the energy planning use case, the absence of comparable baseline practices in current municipal workflows limited evaluation to qualitative assessment of platform capabilities and user feedback. Definitive statements on the quantitative impact of the platform on energy planning therefore remain dependent on further operational deployment and the establishment of measurable reference points. The data structure, ontology extension and service definition for hydroelectric scheduling was established; however, additional work is required for full integration on the platform. Nevertheless, this use case confirms technical feasibility with quantifiable operational benefits, including estimated monthly savings of 16 kCHF/month and between 96-220kCHF/month through peak reduction for the two use cases considered. Adoption in the DIGICITIES platform offers potential scalability to other hydroelectric sites.

Future priorities focus on operational deployment validation, expanding semantic data product catalogs to address data availability constraints, establishing community governance frameworks to manage ontology development and prevent fragmentation, and integrating data sharing agreement middleware for commercial exchange. Strategic priorities include evolving toward decentralized storage architecture enabling user-controlled infrastructure, implementing advanced semantic reasoning over taxonomic hierarchies and external ontology mappings, aligning with emerging data space initiatives including GAIA-X, IDSA and Swiss Data Alliance, and cultivating ecosystem incentives through regulatory mandates and financial mechanisms.

The future success and widespread implementation of the platform depends on demonstratable value from use cases that quantify operational benefits, establishing governance that maintains semantic integrity as usage scales, and creating market conditions that encourage the exchange of valuable data products. Ongoing projects including SWEET CoSi, REFORMERS, and the NEST at Empa, provide opportunities for iterative refinement and additional research into promising new features that could further improve the value proposition.

9 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 project collaborators, stakeholders and knowledge community from the kick-off is shown in Figure 20.

Figure 20: DIGICITIES stakeholder network

10 Acknowledgements

We would like to thank the need-owners and collaborators that have supported this project from the beginning. The generation of text in the report was assisted by AI tools Grammarly and Claude.ai for readability and structure.

11 Publications

1. Allan, J., Mangili, F., Derboni, M., Gisler, L., Hainoun, A., Rizzoli, A., … Sulzer, M. (2023). A semantic data framework to support data-driven demand forecasting. In Journal of physics: conference series: Vol. 2600. Energy & thermal management systems (p. 022001 (6 pp.). https://doi.org/10.1088/1742-6596/2600/2/022001
2. Rubattu, Nicolò, Gabriele Maroni, and Giorgio Corani. "Electricity Load and Peak Forecasting: Feature Engineering, Probabilistic LightGBM and Temporal Hierarchies." International Workshop on Advanced Analytics and Learning on Temporal Data. Cham: Springer Nature Switzerland, 2023.
3. Allan, J., Derboni, M., Acero González, S., Mangili, F., Marchiori, E., Marquant, J., … Rizzoli, A. (2024). The potential of linked data and semantic technologies in data platforms for urban energy planning. In F. Kupzog (Ed.), OVE-Schriftenreihe: Vol. 113. ComForEn 2024. 13. Symposium communications for energy systems (pp. 83-89). Österreichischer Verband für Elektrotechnik.
4. Horak, D., Lorenzen, L., Kurzidim, J., Hainoun, A. (2024). A use case driven design approach of an energy data exchange platform. International Sustainable Energy Conference - Proceedings, 1. https://doi.org/10.52825/isec.v1i.1150
5. Allan, J., Munoz, E., Acero González, S., Bazazzadeh, H., Bellizio, F., Cai, H., … Mavromatidis, G. (2024). A hierarchical knowledge framework for digital twins of buildings and their energy systems (position paper). In J. P. A. Almeida, C. Di Ciccio, & C. Kalloniatis (Eds.), Lecture notes in business information processing: Vol. 521. Advanced information systems engineering workshops. CAiSE 2024 international workshops (pp. 53-58). https://doi.org/10.1007/978-3-031-61003-5_5
6. Koirala, B., Cai, H., Khayatian, F., Muñoz, E., An, J.G., Mutschler, R., Sulzer, M., De Wolf, C. and Orehounig, K., 2024. Digitalization of urban multi-energy systems–Advances in digital twin applications across life-cycle phases. Advances in Applied Energy, 16, p.100196.
7. Hainoun, A., Stortecky, S., Horak, D., Serageldin, A. (2025). Monitoring and evaluation of integrated energy solutions in a Zero-Energy building (ZEB) in Vienna, Energy and Buildings https://doi.org/10.1016/j.enbuild.2025.115906
8. Serageldin, A., Mihály, N., Rezk, K., Hainoun, A. (2025), Development and Validation of Reduced-Order Building Energy Model for an Austrian Zero-Energy School Using IDA ICE. Journal of Building Engineering https://doi.org/10.1016/j.jobe.2025.114401
9. Allan, J., Groen, J., Widl, E. (2025) REFORMERS D5.2: Energy data space for digital twins. (Pending publication)

Project Deliverables:

D1:	Infrastructure and implementation
D2:	Data structures and their application
D3:	Data processing for demand projection
D4:	DIGICITIES usecase learnings and opportunities
D5:	Stakeholder engagement

12 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.

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.

Derboni, M., Salani, M., Rizzoli, A.E., 2021. Flexible Scheduling of Residential Energy Loads, the Optiflex concept.

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

Holles, S., Rütten, F., Türkucar, T., Oliver Thier, C., 2021. Datahub Schweiz. 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.

Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. J. Stat. Softw. 27, 1–22.

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.

Mangili, F., Derboni, M., Zambon, L., Giuffrida, V., & Salani, M. (2026). Enhancing Peak Shaving Efficiency in Small Hydro Power Plants Through Machine Learning-Based Predictive Control. Energies, 19(4), 985. https://doi.org/10.3390/en19040985Mosavi, 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.

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.

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.

Spiliotis, E., Petropoulos, F., Kourentzes, N., Assimakopoulos, V., 2020. Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. Appl. Energy 261, 114339.

Swiss Federal Office of Energy, 2022. Energy Strategy 2050 [WWW Document]. URL https://www.bfe.admin.ch/bfe/en/home/politik/energiestrategie-2050.html (accessed 12.23.22).

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.