Ref. Ares(2023)8218369 - 01/12/2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957047.

H2020-LC-SC3-EE-2020-1/LC-SC3-B4E-6-2020

Big data for buildings

Building Information aGGregation, harmonization and analytics platform

Project Nº 957047

Description of the end-user communication and security layers

Deliverable 3.2

Responsible:	Inetum Realdolmen (Inetum BE)
Document Reference:	D3.2
Dissemination Level:	Public
Version:	1.0
Date:	30/11/2023

Executive Summary

BIGG, or Building Information aGGregation, is a European Commission H2020 project[1] that aims at demonstrating the application of big data technologies and data analytic techniques for the complete buildings life-cycle of more than 4000 buildings in 6 large-scale pilot test-bed. The project will deliver an extensible, open, cloud-based data pipeline consisting of data ingestions, harmonization and analytics for batch and real-time analytics that supports a wide range of services and new business models.

This document presents the components encompassing the communication layer, graphical interface, and security measures for BIGG.

The Communication Layer is responsible for developing the essential infrastructure to ingest and expose a wide range of building data. This layer plays a crucial role in handling raw data, which is ingested and channelled through the BIGG streaming bus system, where it is intercepted, securely stored in dedicated repositories, and harmonized into a consistent format.

The Graphical Interface builds upon this harmonized data, facilitating its publication within the messaging system, storage in dedicated repositories, and synchronous interception by AI toolbox components. These AI-driven tools process data, activate specific pipelines, and extract valuable insights. These insights are then seamlessly integrated into the BIGG system, accessible to dashboard controllers, and formatted for representation in custom dashboards and external systems.

The Security section is an integral component of the BIGG framework, ensuring data integrity, General Data Protection Regulation (GDPR) compliance and confidentiality at every step of the process, from data ingestion to dissemination.

BIGG goes beyond theory by adopting a business case-oriented approach. To meet the diverse needs of various data-based verticals and specialized business scenarios, end-user applications—both graphical and non-graphical—are developed as practical solutions. These applications empower users to harness innovative data-based services, whether by integration into existing systems, the creation of new interfaces, or the development of mobile applications and monitoring tools.

This document provides a comprehensive overview of the BIGG framework, its key Communication layer, Graphical interface and Security measure components, and its capacity to help organizations fully leverage their data while maintaining robust security measures.

[1] https://www.bigg-project.eu/

Contributors Table

Document Section	Author(s)	Other Contributors to results	Reviewer(s)
Introduction	Sampath Mukherjee, Nico Vermeir (Inetum BE)	-	Antony Uhlmann (Helexia) Nicolas Pastorelly (CSTB) Sylvain Rival, Michel Hours (Inetum FR) Stoyan Danov (CIMNE)
Communication Layer	Jarne Kerkaert, Nico Vermeir (Inetum BE) Nicolas Pastorelly (CSTB) Jordi Carbonell (CIMNE) Riccardo Devivo, Romain Hollanders (Helexia - Energis)	Michel Hours (Inetum FR) Jordi Cipriano, Eloi Gabaldon, Stoyan Danov (CIMNE) Josep Maria Granollers, Óscar Sánchez Regueras, Oriol Escursell Jové (ICAEN) Efstratios Keranidis, Georgios Kazdaridis (DomX) Adelfio D'Angiò, Mario De Marco (Intuicy) Oriol Güell, Àlex Almendro (iCAT) Alexiou Konstantinos, Antoniou Iro (CORDIA) A. Giannatos (HERON) Frederic Wauters (Helexia - Energis)	
Graphical applications	Jarne Kerkaert, Nico Vermeir (Inetum BE) Jordi Carbonell (CIMNE) Romain Hollanders (Helexia - Energis) Pierre Lehanneur (Helexia) Polychronis Symeonidis (DomX)		
Security	Dennis Decoster, Veerle Van Keyenbergh, William Debruyn, Sampath Mukherjee (Inetum BE) Frederic Wauters (Energis) Polychronis Symeonidis (DomX) Jordi Carbonell (CIMNE)		
Conclusion	Sampath Mukherjee (Inetum BE)	-	

Table of Contents

List of Figures

Table of Acronyms and Definitions

Acronym	Definition
EC	European Commission
WP	Work Package
EPC	Energy Performance Certificates
CMSS	Computerised Maintenance Management System
GDPR	General Data Protection Regulation
GUI**Graphical User InterfaceAITB**	BIGG AI Toolbox
RAF	Reference Architecture Framework
ISP	Integrated Scheduling Programming
HVAC	Heating, ventilation, and air conditioning
API	Application Programming Interface

I. INTRODUCTION

I.1.1. Purpose and structure of the document

The primary objective of this document is to offer stakeholders, including those who may not be familiar with the intricate details of WP3, a comprehensive understanding of the tasks undertaken and the outcomes achieved. By detailing the purpose and final status of each task within WP3, this document aims to provide transparency and insight into the technical aspects of the project, fostering better collaboration and understanding among project contributors and stakeholders.

Work Package 3 is one of three technical work packages in the BIGG project, next to work packages 4 and 5. Work Package 3 can be considered as the facilitator layer that allows data ingestion and exposure to the different BIGG components to start producing analytical.

Work Package 3 contains three tasks:

- Task 3.1: Communication Layer Development
  - The primary goal of this task is to establish a robust communication layer for the BIGG system. This layer serves as the backbone for data exchange and interaction between the various components, ensuring efficient and effective communication throughout the project lifecycle.
- Task 3.2: Graphical User Interfaces (GUIs) Development
  - This task focuses on the creation of GUIs that enhance the user experience and facilitate user interaction with the BIGG system. These interfaces play a crucial role in making the analytical capabilities of the project accessible to users.
- Task 3.3: Security Layer Development
  - This task elaborates on the implemented security measures, protocols, and the overall effectiveness of the security protocols put in place in protecting the integrity, and availability of data, safeguarding the use cases against potential threats.

This document corresponds to the deliverables of work package 3.

The goal of this document is to detail the three technical tasks defined in Work Package 3, every task details its own purpose, and final status.

The document is structured in the same order as the task definitions except we also provide the information on how to replicate the use cases for anyone external to the project. After this introduction chapter the document describes the Communication Layer (section II), the Graphical Interfaces (section III), Replication of the use cases (section IV) and the Security layer (section V).

I.2. Boundaries between WP3 and WP4

Work Packages 3, 4 and 5 are technical work packages of the BIGG project. Work Package 5 is the development of the AI toolbox (for further details consult D5.2). Work Packages 3 and 4 are closely related so it might be difficult to see where one ends and the other begins.

Work Package 4 is the Data Harmonization Layer (for further details consult D4.2 and D4.3). Its objectives are:

- Analysis of the datasets and the different formats and ontologies to be incorporated into the BIGG solution.
- Development of the BIGG Standard Data Model for Buildings.

- Development of data transformations and mappings to external data sources and ontologies, and connection of the BIGG data model to the use cases tested in the pilots.
- Contribution to the development of the European standards and ontologies.

In short, Work Package 4 will define and create the BIGG data model and the logic to shape incoming and outgoing data into that data model.

Work Package 3 deals with ingesting and receive harmonize data according to the descriptions of the BIGG Data Model created in Work Package 4. The data ingested through the communication layer (Task 3.1) will pass through the harmonization process, the harmonized data will go through the AI Toolbox (WP5, which is the actual user of the data) and be stored or displayed to the user (Task 3.2). All of these steps will happen in a secure way (Task 3.3). Storing data before or after harmonization can optionally be done depending on the use case.

II. Communication Layer

II.1. Introduction

The communication layer was depicted as "responsible for developing the necessary components to **ingest** and **expose** all types of building data". Regarding the state-of-the-art reference architecture framework (RAF) which has been designed in WP2 (see D2.3), the communication layer components are laying in the northbound and in a southbound of the architecture as presented in the following simplified schema:

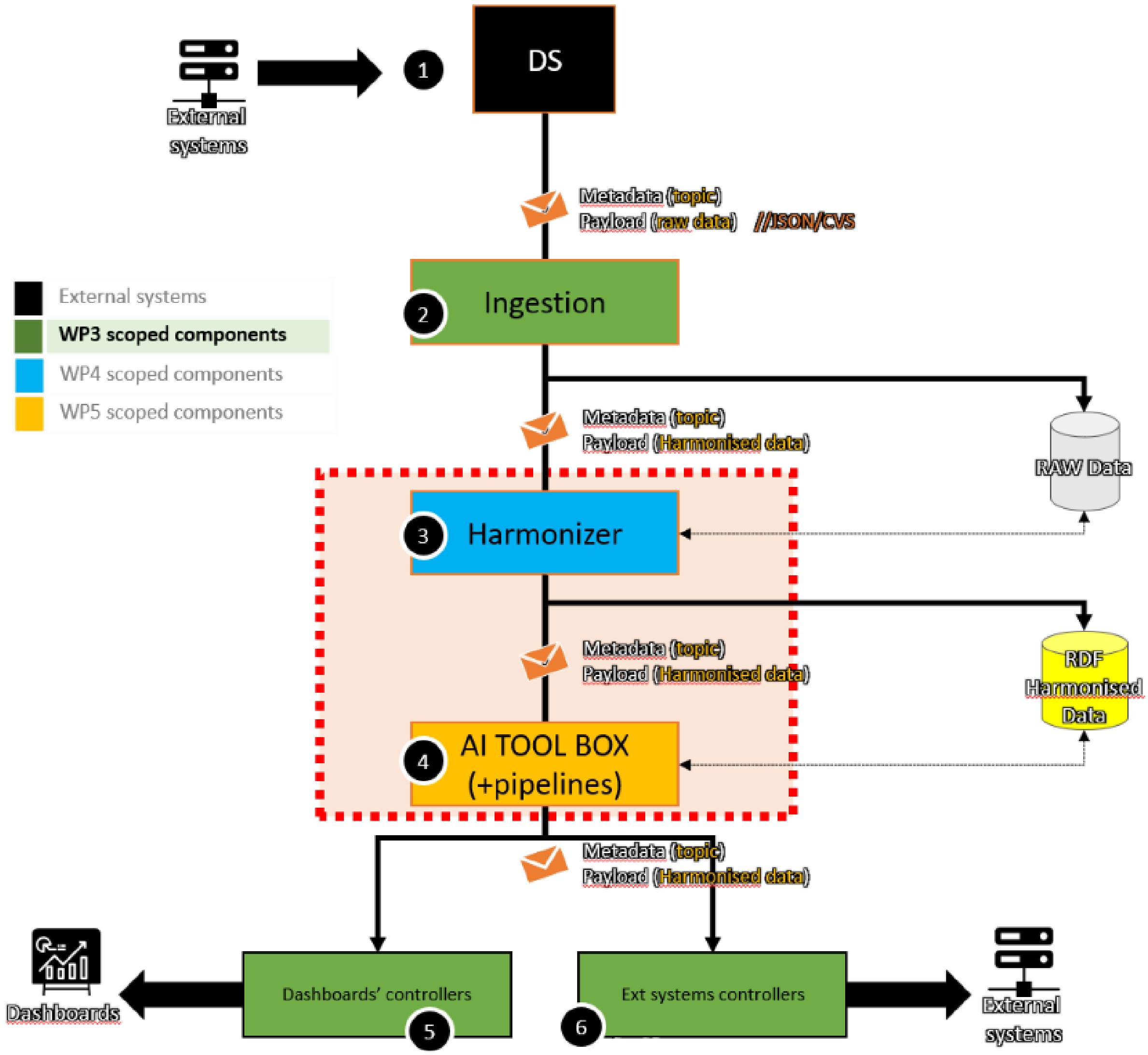


Figure 1 – Simplified communication layer architecture

Next to the Northbound of the BIGG architecture, technical services of external systems are responsible of exposing endpoints to inject datasets (see 1 in **Error! Reference source not found.**) in the BIGG pipelines.

The backbone of the overall BIGG workflow is a messaging system using Kafka to exchange data between the different components. Nevertheless, BIGG components are designed in such a way that, if required, they can be used independently in a less integrated manner (cf. D2.2).

To collect raw data, **Ingestion components** (see 2 in **Error! Reference source not found.**) have been designed to support various protocols (HTTP, MQTT etc.). MQTT's security relies on features like Transport Layer Security (TLS/SSL) for encrypted communication, robust authentication mechanisms, and access control, making it suitable for secure IoT applications. HTTP's security is ensured through TLS/SSL encryption, various authentication methods, and access control, rendering it secure for web applications and data transfer when configured with proper measures.

Ingested raw data get pushed on the BIGG streaming bus system and get intercepted to be stored in raw-data-dedicated stores and to trigger **harmonization components** which transform raw data into harmonized data following WP4 specifications (see 3 in **Error! Reference source not found.**). The harmonized data get published on the messaging system and can be stored in harmonized-data-dedicated stores and/or can be synchronously intercepted by configured AI toolbox components that process it to activate specific pipelines and create knowledge out of it. Insight data created by AI toolbox can then be broadcasted in the BIGG system to be intercepted by **dashboard controllers** formatting this data for representation in specific custom dashboards or **external system controllers** intercepting this data to push it back to specific external systems. These communication southbound components are to be adapted to fit every specific customer requirement (see section III for details).

II.2. Communication Layer implementation in business cases

The introduction above presents the shared architectural vision which has been set inside the consortium during the initial phases of the project. The project is managed in a pragmatic approach where phases of design and phases of implementation are combined in technical sprints. So, every business case has started to incrementally implement the BIGG RAF adapted to its local constraints.

II.2.1. Business cases #1, #2 and #3 – Case Study Area: Catalonia (Spain)

Business cases 1, 2 and 3 has demonstrated the potential of BIGG to be beneficial for both the energy consumers and providers who are participating. Now, in order to provide a scalable solution, BIGG uses large datasets.

These datasets are related to energy consumption or demand in public and residential buildings in the region. These data sets are:

- Dataset 1: A database of 4,000 public buildings of the Government of Catalonia.
- Dataset 2: An existing database of about 1 million residential and tertiary buildings from Energy Performance Certificates (EPC).
- Dataset 3: Monitoring data: Existing database with 272 monitored buildings.
- Dataset 4: Data from the computerised maintenance management system (CMMS)
- Dataset 5: HVAC remote controls: In 40 public buildings, existing database with data from HVAC remote controls, with data from air-conditioning and heating system remote controls.
- Dataset 6: BIM models: Existing BIM models for 25 buildings.
- Dataset 7: Spanish municipal cadastral data in INSPIRE format[2]

[2] https://inspire.ec.europa.eu/data-specifications/2892

Status

The following logical diagram depicts the simplified pipeline to be implemented and its main components:

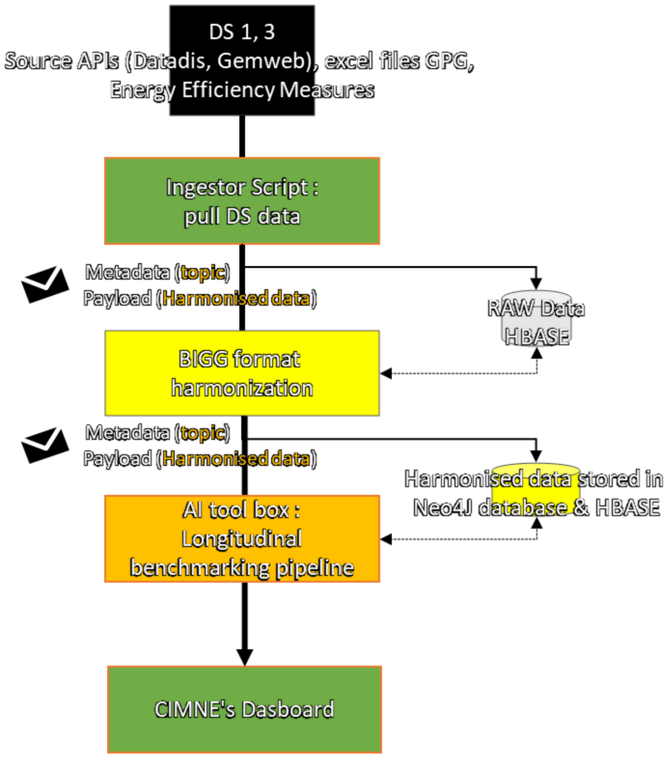


Figure 2 – Communication Layer Architecture

<u>Sources import process</u>

For BC 1, 2 and 3, data are being collected from different available external resources. For data collection, specific services have been implemented for each data source. Depending on the characteristics of the data source we have two categories of collection.

a) Data in existing databases with API service connection.

In this case, the API client is implemented specifically for each resource. The implementation has been done in a flexible way to facilitate the use of these services to new organizations (logging in with the provider's access credentials). These services allow automatic and recurring data requests based on the characteristics of the data.

The sources connected to this service are:

- Electricity consumption. Monthly and hourly electricity consumption of 4000 buildings. Provider: DATADIS[3] Database with all Spain electricity consumptions.
- Gas consumption: Monthly. gas consumption of 1100 utility points of delivery. Provider NEDGIA database with all Catalan Government consumption points.
- Building Performance Certificates (BPC) – Information of more than 1,2 M of BPCs from open data transparency portal of the Catalan Government[4].
- Weather data for all Europe stations. Provider: DARKSKY[5].

[3] https://www.datadis.es/en/data

[4] Certificats d'eficiència energètica d'edificis | Dades obertes de Catalunya (transparenciacatalunya.cat)

[5] https://darksky.net/

- HVAC and Buildings Management Systems (multi-vendor). ICAT has installed a gateway (IXON) in all buildings with BMS with BACnet protocol (272). This allows data collection via VPN from all systems. Provider: IXON[6].
- Billing information systems. Monthly and hourly gas and electricity consumption data. Provider: GENWEB[7].
- ICAT Building Information System. Description information of buildings manage by ICAT (~400 buildings)
- Building Maintenance data. Data of building zones, assets and work orders for more than 300 buildings. Provider: JG MANTEST[8]

b) Manual export files.

In cases where the data is hosted in DB or systems without remote access possibilities, data ingestors have been prepared that allow uploading files in standard format to be processed. For these cases, import formats have been prepared and initial data uploads have been performed. These resources are:

- Descriptive data of the buildings. Provider: GPG. Property data manager of the Government of Catalonia, Department of Economy and Finance
- Energy efficiency measures applied in the buildings of the Government of Catalunya. Provider: ICAEN. Information of EEM applied in Catalan Government buildings (400 registers).
- Cadastral information, in ispire format, on buildings of the Generalitat de Catalunya. Provider: Sede cadastro Español[9].
- Building Performance Certificates (BPC): Processing the single results of each BPC in XML format

Ingestion process

Ingestors are custom-made applications to collect information from different sources. These sources can be found in different formats such as Excel/CSV files, REST APIs or Web forms. The ingested information is channelled through the Kafka[10] broker to different storage services (raw data) or preprocessing (harmonization). Kafka is the main component in the implementation for this part of the process. The implementation of the ingestors in this pilot has been done by integrating them directly into the production code.

It is highlighted that for some sources it is necessary to launch injectors in the MapReduce paradigm to distribute the executions due to the large amount of data and the slowness of the API responses.

Link data and harmonization process

In cases where harmonization-pre-processing is required, the harmonization component collects data and information from the Kafka broker and processes it. This information is harmonized using the harmonizers proposed in BIGG, for some sources, a custom BIGG-derived harmonizer developed by CIMNE is used.

6 https://www.ixon.cloud/

7 http://www.gemweb.es/

8 https://www.testjg.es/manttest

9 CARTOGRAFÍA CATASTRAL - INSPIRE (minhap.gob.es)

10 Apache Kafka

To execute the harmonization process, a custom mapper has been created for each data source. This mapper contains the equivalence between the data in the initial format (raw data) and the data in the format of the BIGG model (harmonized data). The harmonization component collects the ingested data, searches for the source mapper, and performs the transformation of the data.

Data storage

The data in this use case is stored in two phases.

The first one is to store the raw data obtained directly from the external source, before harmonization. This is done to keep a copy of the gathered data in case we need to process it again. These BCs use HBASE[11] to store the data in this phase.

The second is to store the harmonized data, in the final database, to be used by the AI toolbox and Visual interface. We use Neo4j[12] for storing building information data because it is RDF compatible, and HBASE to store timeseries data because it has a very good performance when storing large amounts of data.

Presentation

The presentation of the data in the current version is done through a User Interface developed ad-hoc to cover the BC 1, 2 and 3 user's needs. The main function of the UI is to allow end-users to visualise and manage the ingested and harmonised data (favouring the linking of data between them and the initial exploration) and explore the results of analytics.

The presentation of the results of the AI toolbox has been implemented in CIMNE's Dashboard when required by the use cases. Some examples can be seen in section III.1.2.a.

II.2.2. Business cases #4 and #5 - Case Study Area: Athens (Greece)

This case study is about Interamerican and Vodafone building which consumptions are managed by the company CORDIA. It is structured in two business cases in turn split into three use cases:

- BC 4 - Energy Performance Contract-based savings in commercial buildings
    - UC 8 – Assets management
    - UC 9 – Actual savings tracking
    - UC 10 – Reporting about EPC life cycle
- BC 5 - Buildings for occupants: Comfort case
    - UC 11 - Optimisation using weather forecasts
    - UC 12 - Optimisation using occupancy forecasts
    - UC 13 - Optimisation using price forecasts

In terms of communication, the inputs, computations and outputs are presented below. The implemented modules offer:

- Inputs
    - Data extracted from BMS, IoT sensors (consumptions, temperatures, occupancy, etc.), utility invoices, status and forecast from weather services

[11] https://hbase.apache.org/

[12] https://neo4j.com/

- Computations
    - Historical computation - Identification of best interactions between historical consumption and influencing factors such as weather and occupancy (based on IPMVP protocol [13])
    - Model generation - Identification of baseline models with dependencies from influencing factors, based on IPMVP.
    - IPMVP criteria calculation and comparison to existing methodology
    - KPIs generation - Relevant KPIs which can be viewed in dashboards and reports
    - Optimisation service module addressing the different goals related to energy efficiency, renewable energy usage, occupancy, comfort and cost of the building
- Outputs
    - Baseline consumption model to compare with the actual data, including all the model meta data (cross validation data, model evaluation scores, removed outliers, etc.)
    - Optimization parameters for the rule engine of the Controller module to pilot the building appliances (HVAC)
    - Dashboards and reports

Status

The following logical diagram depicts the simplified pipeline implemented and its main components:

[13] https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp

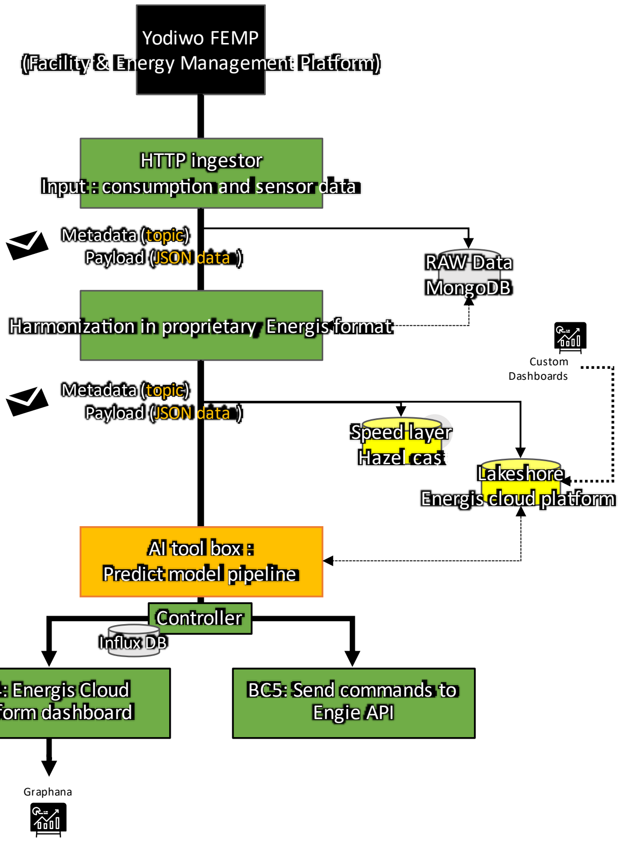


Figure 3 – Ingestor Pipeline

- **Ingestion:** the "solicited-ingestor" microservice retrieves consumption data from the Yodiwo FEMP platform in a solicited way through active REST API invocation and produces messages on a dedicated Kafka topic
- **Datalake:** the "mongodb-adapter" microservice retrieves the raw data messages from Kafka and stores raw data into the MongoDB repository
- **Harmonization:** the "processor" microservice retrieves the raw data messages from Kafka and transforms raw data messages into an internal JSON format called "processed message"
- **Speed layer:** the "memory-adapter" microservice retrieves all the transformed messages and stores its content into an In Memory Data Grid implemented by Hazelcast
- **Lakeshore:** the "adapter" microservice sends all the transformed messages to the Energis.cloud platform to store them in the Cassandra lakeshore. The energis.cloud platform offers a GUI to show dashboards and to send reports to the final user. This GUI is presented in section **Error! Reference source not found.**- **Error! Reference source not found.** below.
- **AI Toolbox (AITB):** manually launching a script, this set of AI and ML tools retrieves metrics and historical weather data from Energis.cloud invoking its REST API services, then the AITB elaborates a consumption regression model whose formula is then incorporated to a metric in Energis.Cloud. This output is then used to compute all the required KPIs for BC4, which are in turn incorporated in dashboards and reports in Energis.Cloud.
- **Controller:** the "controller" module receives parameters from the AITB module, retrieves weather status and forecasts (temperature, irradiance, etc.) from the WeatherBit service and receives appliances status from Cordia Connect service to tune its internal rule engine and, in turn through Cordia Connect, sends control commands to all the HVAC assuring both consumptions reduction and comfort levels for the

occupant. The "controller" has its own InfluxDB repository and uses Grafana to store and display relevant information on its status for debugging and monitoring purposes as presented below.

The following diagram shows in a more technically detailed way the status of BC4 and BC5 implementation:

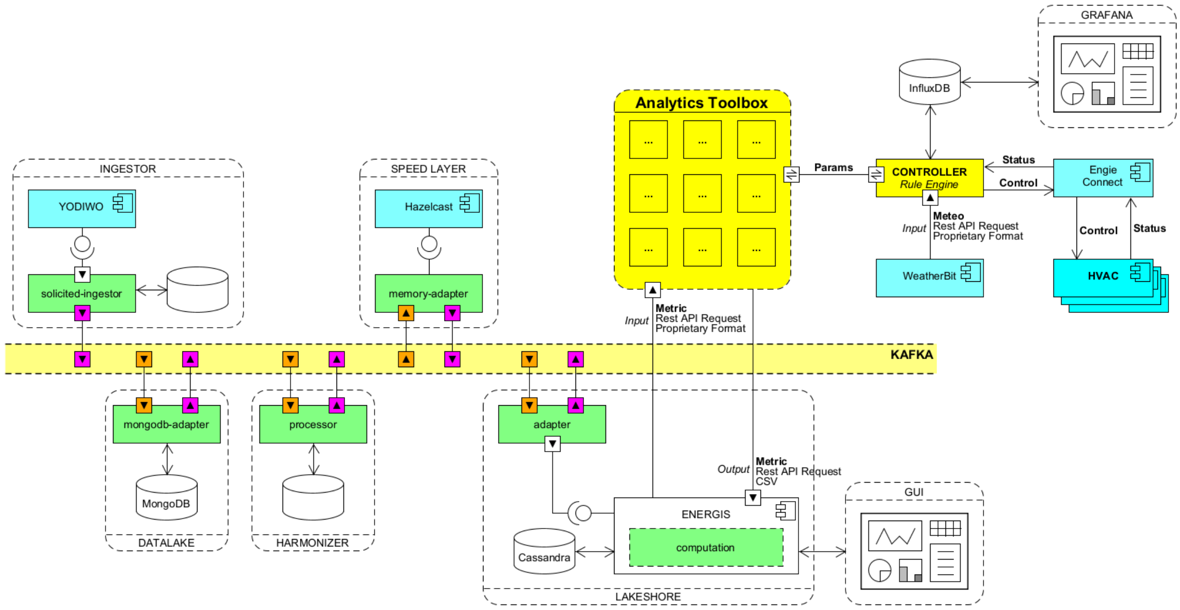


Figure 4 – BIGG Architecture Diagram

This is the list of activities that were completed to fulfil the requirements exposed in the beginning part of this section:

- Harmonization
  All the different components had to use the BIGG RDF ontology format instead of proprietary JSON format. This format is standardised and structured and it was necessary to adapt the Energis.Cloud platform to comply with it. In particular, to build the solutions for BC4 and BC5, it was necessary to be able to adapt and extend the allowed asset types (e.g. building, equipment, Energy Performance Contract main scope, Energy Performance Contract sub-scope, etc.) and their inherent properties. This was done by introducing a new "asset role" concept to Energis.Cloud together with a concept of "meta-metrics"[14] allowing us to freely define new asset types and, for each new type, to freely define their properties and how they need to manipulate data based on these properties. Thanks to these changes, it became possible to smoothly interface Energis.Cloud with the BIGG ontology and to ensure resilience against future adaptations.

- Integration
  The Analytics Toolbox is integrated with the platform architecture in terms of communication (e.g., using Kafka topics instead of direct REST API calls) and is wrapped by a microservice exposing its main functionalities. The "memory-adapter" microservice is used by the other modules requiring fresh data in near real-time, moreover it can be used as a common cache to store relevant entities and metrics shared between all the platform components.

[14] The Asset roles and Meta-metrics concepts have been later relabeled into Metric groups and Metric definitions, with an additional concept of Metric templates gluing the two concepts. However, for past reference purpose, we included the original names in the text. More documentation about these concepts can be provided on request.

- Persistence
  The library used by the AITB (MLflow) offers the possibility to persist the predictive model as a serialized stream of bytes. Exploiting a dedicated Kafka topic, this model can be persisted into MongoDB as a document to be in a second moment retrieved to speed up AI and ML computation procedures.
- Implementation
  The WeatherBit service is accessed by Energis.Cloud too, so it was cleaner to implement a unique microservice accessing WeatherBit to persist and offer temperature, irradiation and forecast data to all the modules requiring it. The occupancy and pricing data can also be collected by dedicated microservices decoupling the datasources logic from the computational logic. Besides the mentioned “controller” module, in the RAF (Reference Architecture Framework) were described several components that were implemented for modularity and portability sake: the “commander” to read command and propagate results and the “api gateway” to simply access all the different API’s and functionalities offered by the Analytics Toolbox.

All these steps are summarized in the following architectural diagram:

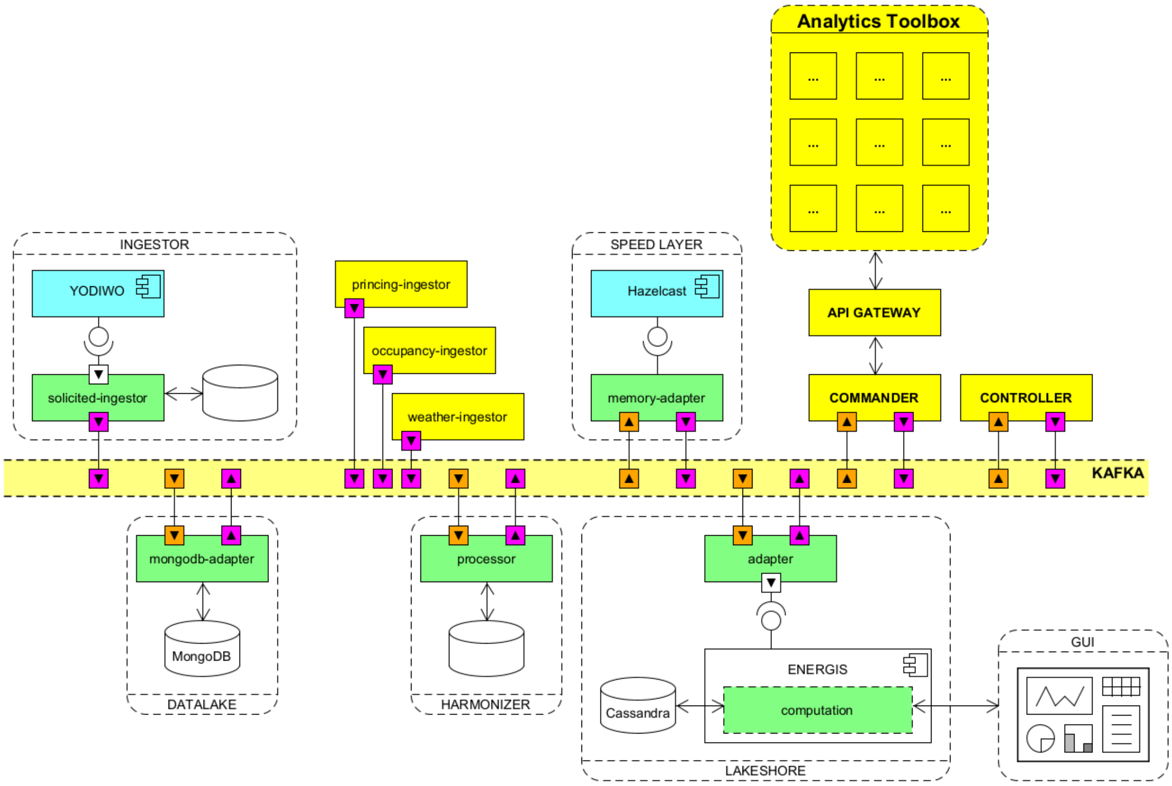


Figure 5 – BIGG Architecture Diagram

II.2.3. Business case #6 - Case Study Area: Several cities (Greece)

Business case #6 will demonstrate the application of Demand Response management on top of electricity and gas consumers of HERON including the gathering and harmonization of data stored in different databases. These data sets include:

- Dataset 1 (HERON): A database including data collected for years for residential and commercial buildings of HERON consumers (subset of 150.000): Aggregated Consumption data Load profiles.
- Dataset 2 (HERON): A database including data collected for 1 year for residential and commercial buildings of HERON electricity consumers:
    - Real-time Consumption and Billing data from 200 consumers
    - Real-time Production and Billing data from up to 20 consumers
- Dataset 3 (HERON): A dataset that includes market and system (RES shares, Hydro power, load forecasts) data from Green Market Operator (EnEx) and TSO (IPTO), curated and maintained by HERON.
- Dataset 4 (domX): A database including data collected for 1 year for residential and commercial buildings of domX clients, characterizing the use of heating and hot water preparation gas boilers (subset of 100): Indoor/outdoor temperature, temperature setpoint, heating water temperature, domestic hot water temperature, heating/hot water usage, boiler modulation / gas consumption. All above parameters reported at minimum per minute.
- Dataset 5 (Building EPC data): EPC data, characterizing building details (characteristics, consumption, efficiency, devices, etc.) is be made available for offline access, to assist the building performance analysis.

The above mentioned datasets reside in two different Cloud Platforms (HERON's and domX's respectively). Each cloud platform has different characteristics and extension capabilities but share some common tools and frameworks.

Status

UC#14 and UC#15

The following flow diagram describes the final status and main characteristics of the components used in UC#15. The infrastructure and components employed in UC#14 follow a similar deployment approach with UC#15, so we refrain from presenting them both here.

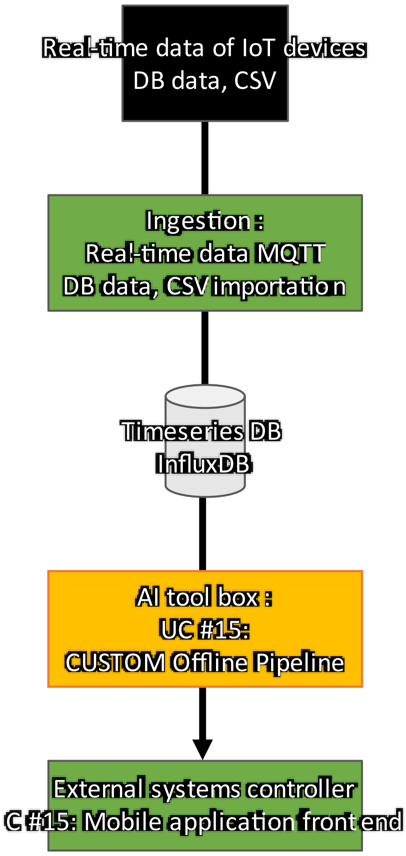


Figure 6 – Components used in UC15

Use Case #15 is ingesting data from Datasets 3 and 4 that are populated from real-time (IoT Sensors) and static data sources in the form of CSVs.

Ingestion:

Real-time data are captured and transferred through MQTT from heating controller elements installed in residential buildings, periodically with a retrieval rate of approximately 10 seconds. The data are then ingested from a microservice in the form of JSON document and transformed to a format ready to be inserted to the TimeSeries Database (InfluxDB).

Static data (EPC, and building characteristic data) are ingested periodically by processing uploaded CSV formatted documents by a service that runs periodically at the start of each day and processes data that are then inserted to a Relational Data Base structure (PostgresDB).

Datalake:

InfluxDB is used to store data captured from real-time IoT sensors. The data is stored raw using the InfluxDB line protocol format.

Harmonization:

Harmonization of data is not required in the initial version of the UC#15 Setup since the Real-time IoT Sensors are built in-house and the data have been adapted from the source to serve the needs of the use case.

AI Pipeline:

After the Ingestion process, we trigger asynchronously and manually an existing AI pipeline that periodically processes the data and provides insights and profiling data of the installation sites. This AI Pipeline performs timestamp alignment, outlier detection, it manages missing data, detects and zeroes-out disconnection windows that are bound to happen with real-time devices deployed in the field.

External System Controller:

DomX's mobile application is used as an External System Controller. The controller receives data from the AI Pipeline which are then presented to the application in the form of a monthly report where users can receive insights about gas consumption usage and consumer behaviour patterns. More details can be found in section 0- DomX User Interfaces.

Business Case 6 implementations are aimed towards preparing and transforming the current architecture to the reference architecture provided by other Work Packages of the project. The main goal is to utilize BIGG components developed for the scope of the project and incorporate them into the existing architecture seamlessly in order to be compliant with the requirements of the BIGG Project.

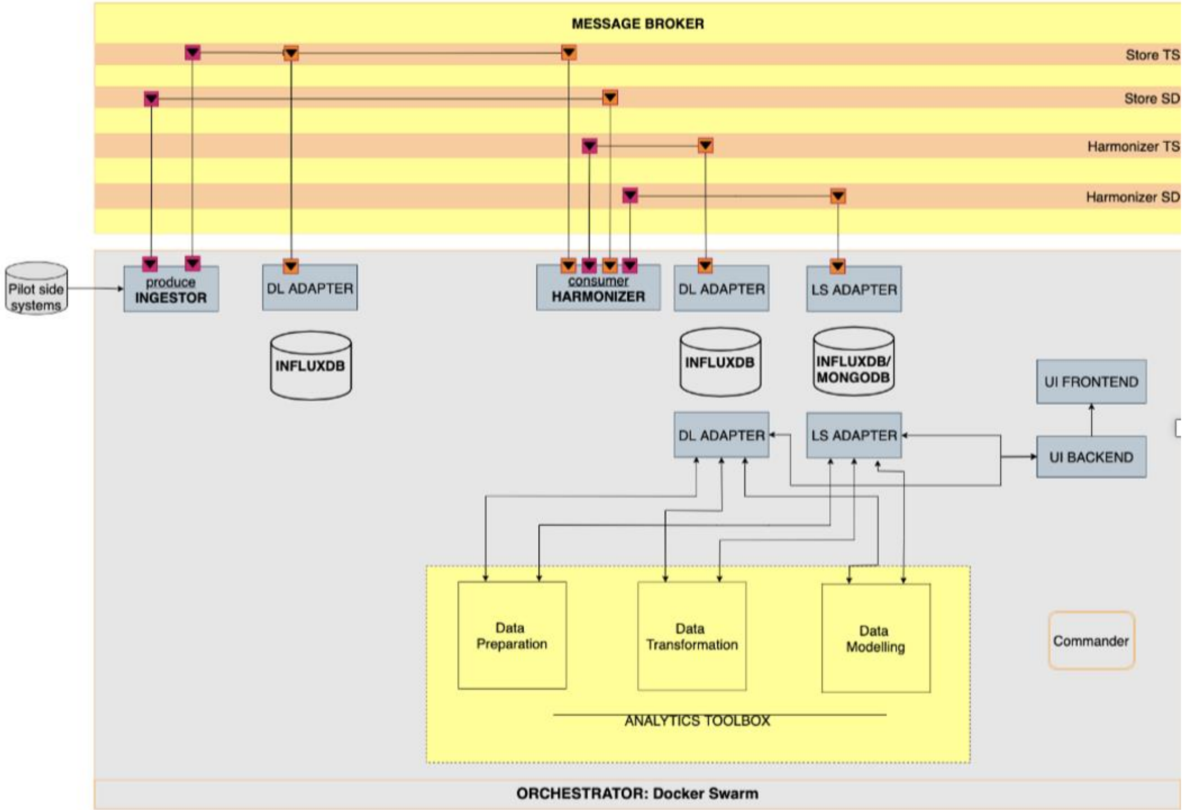


Figure 7 – BIGG Architecture

More specifically we break down the improvements made to the individual components below:

Apache Kafka:

Since most of the BIGG Components implemented use Apache Kafka as a message broker between components we have incorporated Kafka as a framework in the existing architecture.

Ingestion:

The current data Ingestor is be modified to stream events to Apache Kafka topics. Kafka Connect plugins is incorporated to produce raw data to Kafka Topics that is readily consumed by the BIGG Harmonizer component.

Harmonization:

The Data Harmonization is a big part of the BIGG architecture, and we initially planned to use the BIGG Harmonizer component implemented which transforms raw data collected from the data-sources involved in BC#6 and output data in the BIGG Format ready to be used by other components of the pipeline. However, due to business case need that component has not been used for UC 14.

AI Pipeline:

The current AI Pipeline includes the BIGG AI Toolbox developed in WP5. The BIGG AI Pipeline is a big upgrade compared to the existing AI Pipeline, it enhances several aspects of the Data Transformation and integrates Data Modelling. (Readers can consult D5.2 for more technical aspects of AI Toolbox).

It prepares the data (timestamp alignment, outlier detection, missing data management) and it applies steady-state feature transformation. Moreover, in the data modelling aspect it performs time-series analysis, tunes, trains and evaluates the model. The goal was that the BIGG AI Pipeline receives harmonized data through KAFKA topics and perform analytics and produce meaningful insights in an automated way.

External System Controller:

Improvements has been made to the External System Controllers. DomX's Mobile App is updated to conform with the needs of UC#15, by providing a way for end-consumers to participate in gas flexibility services, while balancing the impact of the defined climate comfort. In addition, a new External System Controller has been introduced in the form of a web-based Energy Supplier Dashboard. This Dashboard aims to provide insights and control capabilities to the Natural Gas Suppliers, in order to correct their daily Natural Gas demand predictions and minimize the deviations from the daily supply estimations. We will explain in detail in Section III about the graphical interfaces.

III. Graphical Interfaces

III.1. Status

A key characteristic of the BIGG solution is its modularity. It guarantees a high rate of general reusability. To ensure the fulfilment of the business case-oriented approach of BIGG, end-user applications (both graphical and non-graphical) were developed as solutions that can be employed. In some cases, for multiple diverse data-based vertical scenarios, while in some other cases for specialized business scenarios. These applications allow end-users to capitalize on innovative data-based services, either integrating them into their existing systems so data can be exposed via new dashboards, views, reports (defining APIs, web services or other types of interfaces) or allowing the creation of new ones (e.g. Mobile apps, monitoring tools). To this end for Task 3.2 it was decided to develop:

- A common Pilot Progress Monitoring Dashboard End-user applications and tools developed by each partner separately.

III.1.1. Pilot Monitoring Dashboard

DomX designed and implemented a common dashboard for Pilot monitoring based on the Grafana Framework. This Monitoring Tool is able to receive data from external data sources (pilot-sites) employing various communication channels and data formats.

The implementation is hosted by domX and is based on Grafana. Grafana is a multi-platform open-source analytics and interactive visualization web application. It provides charts, graphs, and alerts for the web when connected to supported data sources. DomX designed and implemented an architecture able to adapt to every Pilot site needs and configurations.

Architecture & Implementation

The Dashboard is able to monitor the Pilot Progress through a set of defined KPIs and visualizes High Level Metrics (Number of Buildings, number of Devices, number of users, etc.) tailored to each Pilot Site, Business Case and/or Use Case.

The architecture is simple, and its core relies on Open-source software Tools. The main component of the software stack is Telegraf, which is an open-source server agent able to collect metrics from a wide array of inputs and write them into a wide array of outputs. It is plugin-driven for both collection and output of data so it is easily extendable.

For the BIGG Pilot Monitoring Dashboard we are utilizing the HTTP API Plugin which creates configurable API endpoints that can ingest different formats of data (JSON, CSV, etc.). We created endpoints for each Pilot that transfer data over HTTPS and are protected with Basic Authentication. These endpoints are then tailored to the data of each Pilot and receive data through HTTP POST requests.

Telegraf receives the data and transforms them into a format that can be stored in the Data Base. The storage structure that we are using is InfluxDB, which is an open-source Time Series Database. The KPI data transmitted to the server through Telegraf are stored in timeseries format which facilitates visualization of data and provides the capability to process that data on the fly by creating complex queries.

Data are then fed to the Grafana dashboard by using the Grafana InfluxDB Datasource plugin, that can natively query and visualize data in the TimeSeries Database.

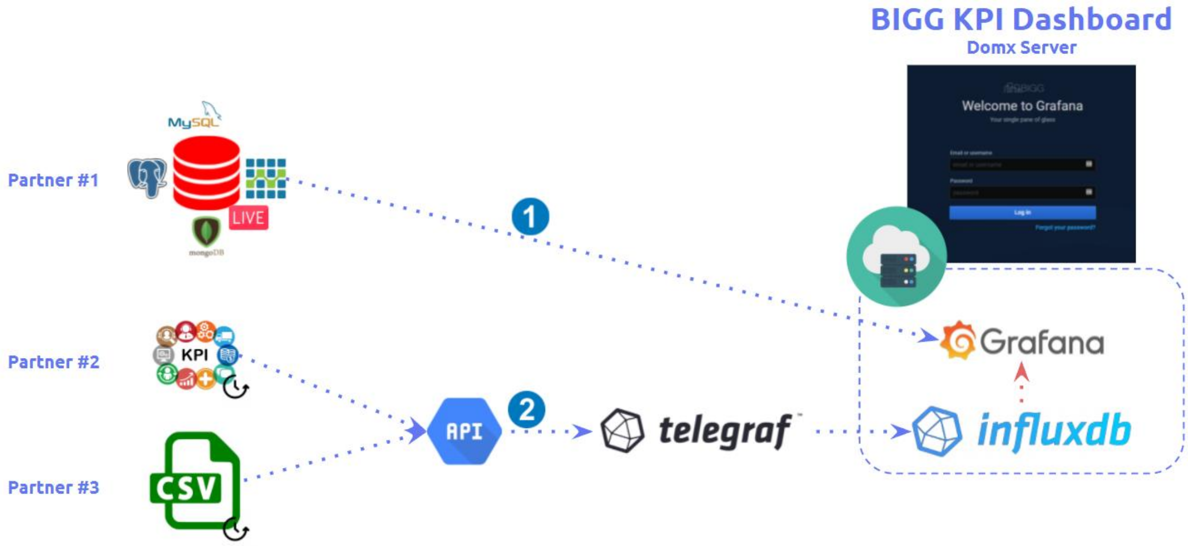


Figure 8 – BIGG Pilot Monitoring Dashboard Architecture

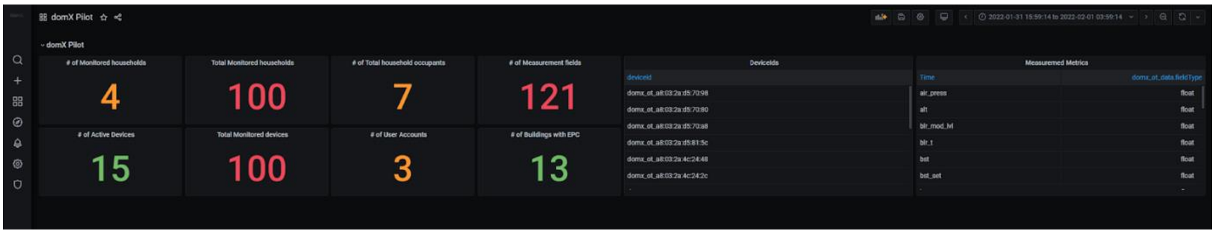


Figure 9 – Proof-of-Concept Implementation for the BIGG Pilot Monitoring Dashboard (Grafana)

For the scope of T3.2 Pilot leaders have discussed about the creation of meaningful metrics and KPIs that can be transmitted to the Pilot Monitoring Platform, to showcase the progress of each pilot and the BIGG Project as a whole.

The initial set of metrics from each pilot are presented below:

BC 1 – BC 2 ICAEN

Pilot Progress KPIs - BC1 - BC2				
ICAEN				
The Common KPI Dashboard will act as an internal tool developed to monitor and showcase the Pilot Progress. Please copy the template to a new Sheet and add below the Metrics/KPIs that are relevant to your Pilot, and will be able to illustrate that progress.				
Name	Type	Unit	Period (Calculation/Transmission to the platform)	Description
Number of Public Buildings (with data)	Variable	Number	weekly	Public Buildings with their data associated to Platform: cadastral data, building location (weather data), link
Target buildings	Variable	Number	weekly	Target Number of Buildings
Availability of monthly electricity energy data (electricity)	Variable	Ratio	weekly	Public Buildings with their monthly electricity data (recorded) associated to Platform
Av. of hourly consumption data (electricity)	Variable	Ratio	weekly	Public Buildings with their hourly electricity data associated to Platform
Av. of Monthly gas energy data (GN)	Variable	Ratio	weekly	Public Buildings with their monthly gas data associated to Platform
Number of buildings with a base line	Variable	Number	weekly	Public Buildings with their baseline consumption calculated by the platform
Users registered	Variable	Number	weekly	Total number of users registered
Active users	Variable	Number	weekly	Share of active users (that at least log in once in 6 months)
Nº of improvement actions collected	Variable	Number	weekly	Amount of improvements in general in a given building or facility.
Share of actions without finish date	Variable	Ratio	weekly	When logging data the finish date of the action is missing, or it is not complete
Share of actions without the value share of EEM over whole building	Variable	Ratio	weekly	When logging data the affected share over the total building of the action is missing, or it is not complete
Number of actions without typology of action	Variable	Number	weekly	When logging data the typology of the action is missing, or lacking in specific details
Number of actions without investment cost	Variable	Number	weekly	When logging data the investment cost is missing
Share of EEM which ROI, Pay-back time and IRR can be calculated based on investment	Variable	Ratio	weekly	Share of EEM which ROI, Pay-back time and IRR can be calculated based on investment
Georeferenced BPC files using UTM	Variable	Number	weekly	Number of BPC files that have been georeferenced using UTM coordinates.
Share of BPC files with cadastral information	Variable	Ratio	weekly	Share of BPC files with cadastral information
Share of BPC data Standardized in INSPIRE format	Variable	Ratio	weekly	Share of BPC data Standardized in INSPIRE format
Number of certificates	Variable	Number	weekly	Amount of certificates which data has been uploaded
Share of certificates with any indicator mapped	Variable	Ratio	weekly	Number of certificates with enough data to map at least one indicator of Level(s)
Share of indicators that have a proposal of future mapping	Variable	Ratio	weekly	Share of indicators that a mapping route has been described for

BC 4 – BC 5 CORDIA

Pilot Progress KPIs				
CORDIA / HELEXIA				
The Common KPI Dashboard will act as an internal tool developed to monitor and showcase the Pilot Progress Please copy the template to a new Sheet and add below the Metrics/KPIs that are relevant to your Pilot, and will be able to illustrate that progress.				
Name	Type	Unit	Period (Calculation/Transmission to the platform)	Description
Target number of EPCs	Constant/Static	Number	once	The Projected number of EPCs that need to participate to the BIGG project from the pilots
Current number of EPCs	Variable	Number	monthly	The number of EPCs currently participating in the BIGG Project from the pilots
Target number of Time series data collected	Constant/Static	Number	once	The Projected number of Time series data collected from the pilots
Current number of Time series data collected	Variable	Number	monthly	The number of Time series data currently collected from the pilots
Target number of EPC models identified	Constant/Static	Number	once	The Projected number of regression models identified in the context of EPC follow up in the pilots
Current number of EPC models identified	Variable	Number	monthly	The number of regression models currently identified in the context of EPC follow up in the pilots
Target average precision of models (CVRMSE)	Constant/Static	Number	once	The average targeted CVRMSE (normalized error) across all identified models
Current average precision of models (CVRMSE)	Variable	Number	monthly	The average achieved CVRMSE (normalized error) across all identified models
Target average bias of models (NMBE)	Constant/Static	Number	once	The average targeted NMBE (normalized bias) across all identified models
Current average bias of models (NMBE)	Variable	Number	monthly	The average achieved NMBE (normalized bias) across all identified models
Target average prediction accuracy of models (R2)	Constant/Static	Number	once	The average targeted R2 (prediction accuracy) across all identified models
Current average prediction accuracy of models (R2)	Variable	Number	monthly	The average achieved R2 (prediction accuracy) across all identified models
Target number of Actuators data collected	Constant/Static	Number	once	The Projected number of Actuators data collected from the pilots
Current number of Actuators data collected	Variable	Number	monthly	The number of Actuators data currently collected from the pilots
Target number of Actuators controlled	Constant/Static	Number	once	The Projected number of Actuators data controlled during the pilots
Current number of Actuators controlled	Variable	Number	monthly	The number of Actuators data currently controlled during the pilots

BC 6 – UC15 DOMX

Pilot Progress KPIs				
DOMX				
The Common KPI Dashboard will act as an internal tool developed to monitor and showcase the Pilot Progress Please copy the template to a new Sheet and add below the Metrics/KPIs that are relevant to your Pilot, and will be able to illustrate that progress.				
Name	Type	Unit	Period (Calculation/Transmission to the platform)	Description
Target households	Constant/Static	Number	weekly	The Projected number of Households that need to participate to the BIGG project from the pilot
Current number of Households	Variable	Number	weekly	The number of Households currently participating in the BIGG Project
AVG Number of occupants/household	Variable	Number	weekly	The average number of households occupants
Total Number of occupants	Variable	Number	weekly	The total number of occupants participating in the domX pilot
Active Devices	Variable	Number	weekly	The total number of Active devices in the domX pilot
Total Number of Devices	Constant/Static	Number	weekly	The target of the total number of Devices in the domX Pilot
Number of User Accounts	Variable	Number	weekly	The total number of Users that Are using the domX Mobile Application
Active Users	Variable	Number	weekly	The total number of Active users that have logged in the domX Mobile Application the Past week
Facility Managers Registered	Variable	Number	Weekly	The total number of Facility manager accounts created (Web-based UI)
Number of Buildings with EPC	Variable	Number	Weekly	The total number of Buildings in the Pilot that have EPC Data
Avg Building Sq. Meters	Variable	Number	Weekly	The average number of sq. meters of the buildings in the Pilot

BC 6– UC14 Heron

Pilot Progress KPIs				
HERON				
The Common KPI Dashboard will act as an internal tool developed to monitor and showcase the Pilot Progress Please copy the template to a new Sheet and add below the Metrics/KPIs that are relevant to your Pilot, and will be able to illustrate that progress.				
Name	Type	Unit	Period (Calculation/Transmission to the platform)	Description
Target households	Constant/Static	Number	weekly	The Projected number of Households that need to participate to the BIGG project from the pilot
Current number of Households	Variable	Number	weekly	The number of Households currently participating in the BIGG Project
AVG Number of occupants/household	Variable	Number	weekly	The average number of households occupants
Total Number of occupants	Variable	Number	weekly	The total number of occupants participating in HERON pilot
Active Devices	Variable	Number	weekly	The total number of Active devices in the HERON pilot
Total Number of Devices	Constant/Static	Number	weekly	The target of the total number of Devices in the HERON Pilot
Number of User Accounts	Variable	Number	weekly	The total number of Users that are using HERON Mobile Application
Active Users	Variable	Number	weekly	The total number of Active users that have logged in the HERON Mobile Application the Past week
AVG number of users without a disconnect of more than 24 hours	Variable	Number	weekly	AVG number of users who have stayed online per day
Number of households with 1 smart plug	Variable	Number	weekly	The number of households that have 1 smart plug installed
Number of households with 2 smart plugs	Variable	Number	weekly	The number of households that have 2 smart plugs installed
Number of households with a flexbible asset	Variable	Number	weekly	The number of households that have a flexible asset
Target households with at least 1 smart plugs	Constant/Static	Number	weekly	The projected number of households that have at least 1 smart plugs installed
Target households with a flexbible asset	Constant/Static	Number	weekly	The projected number of households that have a flexible asset
AVG number of users who have followed 1 DR advice at least	Variable	Number	weekly	AVG number of users who followed at least 1 DR advice per day

Even though the implementation of the Proof of concept of the common visualization platform was successful it was decided by all partners to proceed with using separate tools for illustrating the Pilot progress and KPI data. The decision was taken due to the difficulty of implementing a service to offload data periodically on the BIGG Pilot Monitoring Dashboard, technical security concerns of the process and the heterogeneity of KPI data.

III.1.2. End-user applications & dashboards

CIMNE User Interfaces

A new web application has been developed for the users of business cases 1, 2 and 3.

The development of the user interface serves a dual purpose. Firstly, it enables the efficient management and presentation of data collected across all use cases. Secondly, it facilitates the linkage of data to building entities that may not have been automatically linked, allowing for a more comprehensive data representation. Additionally, the interface streamlines the creation, modification, and correction of input data, ensuring accuracy and completeness in the context of big data analytics.. On the other hand, it allows the visualisation of the results of the data analysis, carried out with the AI toolbox of BIGG, for the use cases that require it.

The main features of the visualisation application are presented below.

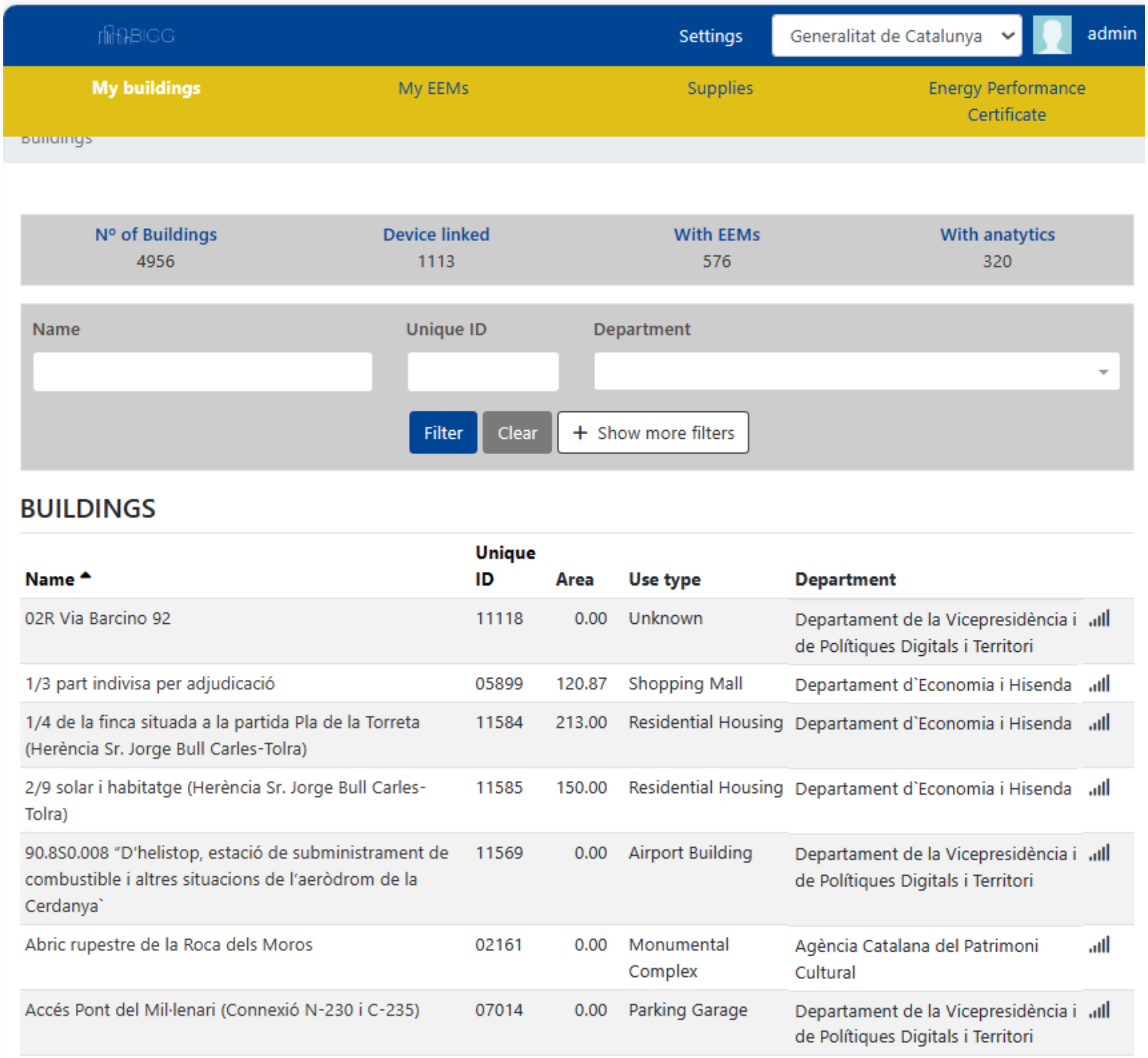
Use case 1: Benchmarking and energy efficiency tracking in public building

The benchmarking of buildings is made using a set of multi-dimensional Key Performance Indicators (KPIs) related with energy use, cost, and emissions. This is referred to as A1 application[15].

List views have been created to facilitate the monitoring of Buildings (>4900) and Devices (> 7000), allowing the search and filtering of them. Each of these lists allows the creation and editing of the content of each individual element, allowing the quality of the input data improvement.

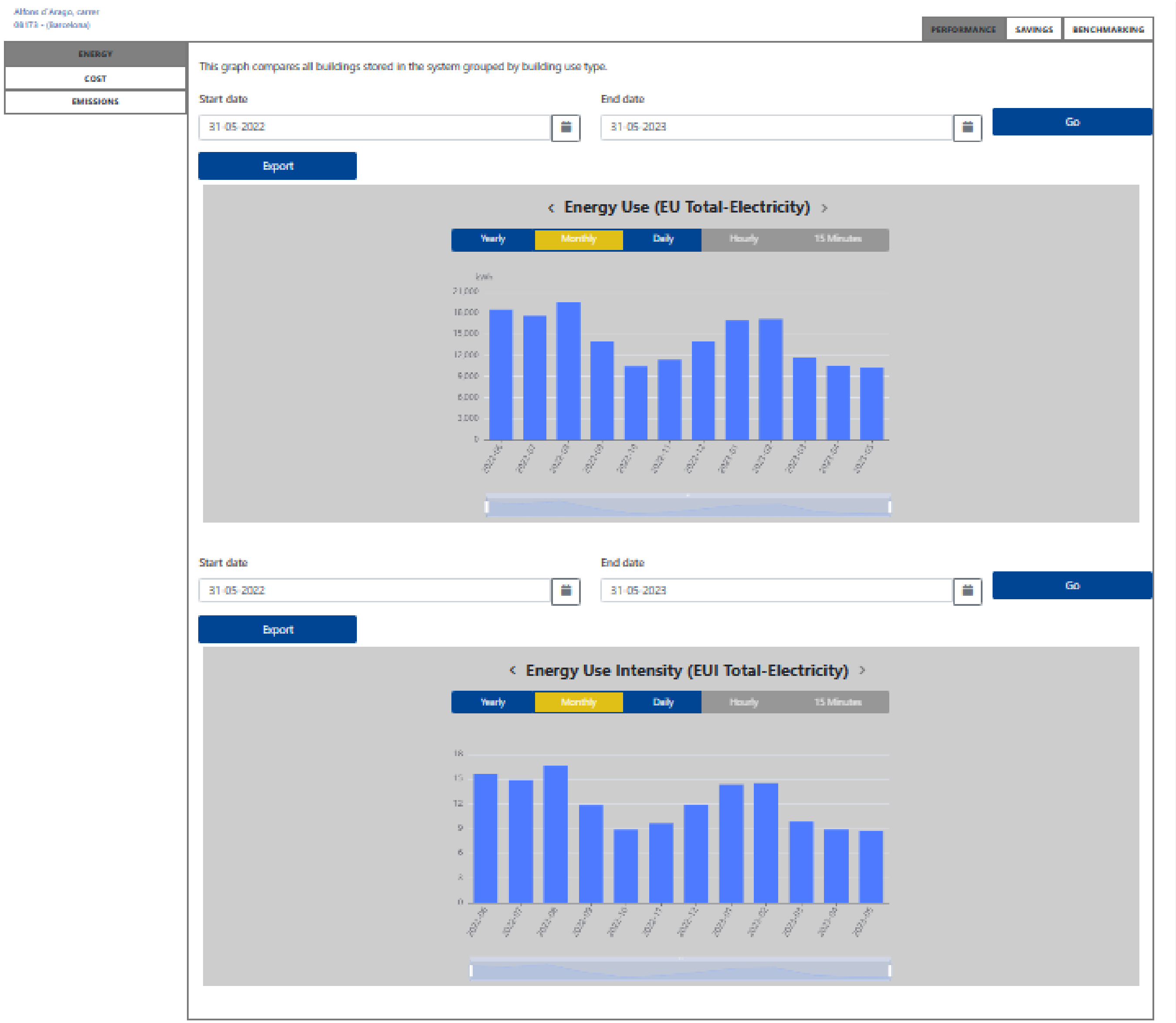
In addition, where the necessary input data is available, analytical views related to the A1 application of the BIGG AI ToolBox have been implemented (for more information see deliverable 5.2 of the project). Below some representative screens of the User interface implementation are presented. a) Buildings List; b) Individual building form; c) individual performance KPIs, d) Similar building comparison.

[15] https://github.com/biggproject/A1-Benchmarking



a) User interface Buildings List

b) User interface of individual building form



c) individual performance KPIs

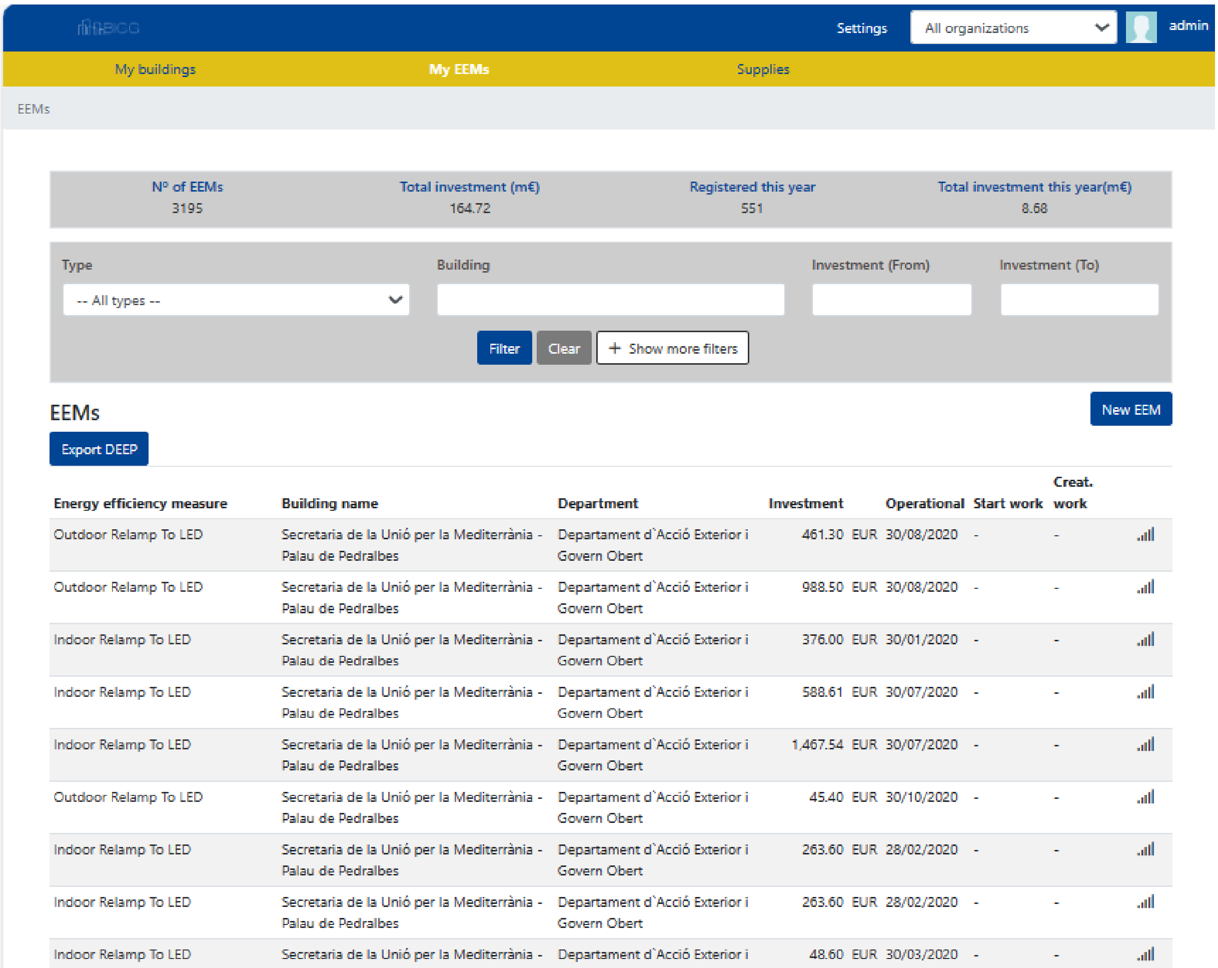
d) Similar building comparison.

Use case 2: Energy Efficiency Measures (EEM) registration and evaluation

This use case, referred to as A2 application[16], uses a Measurement & Verification technique to assess Energy Efficiency Measures (EEMs) in buildings. In addition to the views created for the previous use case, a list view has been implemented to explore the EEMs registered in the application (>3000). Also, a form for the creation of new EEMs, allowing the continuous registration of EEMs, as well as their edition. We emphasize that the selection of measures is based on the taxonomy of EEMs created in the BIGG ontology, thus allowing a greater homogeneity in the registration and facilitating its subsequent processing.

The results view of the AI Toolbox analytics, related to the A2 application, have been created (for more information see deliverable 5.2 of the project).

The following figures show some representative screen shots of the use case:



a) List of measures

16 https://github.com/biggproject/A2-EEM-assessment

Improvement measure

EEM Types

- HVAC And Hot Water Measure
-- Hot Water System Measure
--- Hot Water Final Elements Measure
---- Hot Water Tap Saving Device Installation

Type description

Hot Water Tap Saving Device Installation

Economic investment	Investment Currency
50,000.00	EUR
Start work date	**Start operational date**
-	2020-11-30

Created date

2022-12-05

% of element or zone affected

100

Comments and notes

Projecte conjunt ESE: MCE9 - Instal·lació de carxofes de baix consum per dutxes

b) form for creating individual measure buildings


Start date
End date
Go
Export
Energy Use Savings (EUS-Baseload-)
Start date
End date
Go
Export
Energy Use Savings (EUS-Baseload-Total Consumption)
Start date
End date
Go
Export
Energy Use Savings (EUS-Cooling-)
Start date
End date
Go
Export
Energy Use Savings (EUS-Cooling-Total Consumption)

c) EEMs KPIs

Use Case 3: Integration of INSPIRE spatial data with Energy Performance Certification (EPC)

In this use case, the Energy Efficiency Certificates of buildings have been integrated with INSPIRE geospatial data. For the demonstration of this integration, a demo based on open-source OpenStreetMap WKT (https://gitghub.com/clydedacruz/openstreetmap-wkt-playground) has been created. The following image shows an example of the possible application of the data link, we can see how we can calculate the distribution of the KPI Energy Intensity of buildings with GCP with a geographic selection of the points to be included by means of a geolocated map.

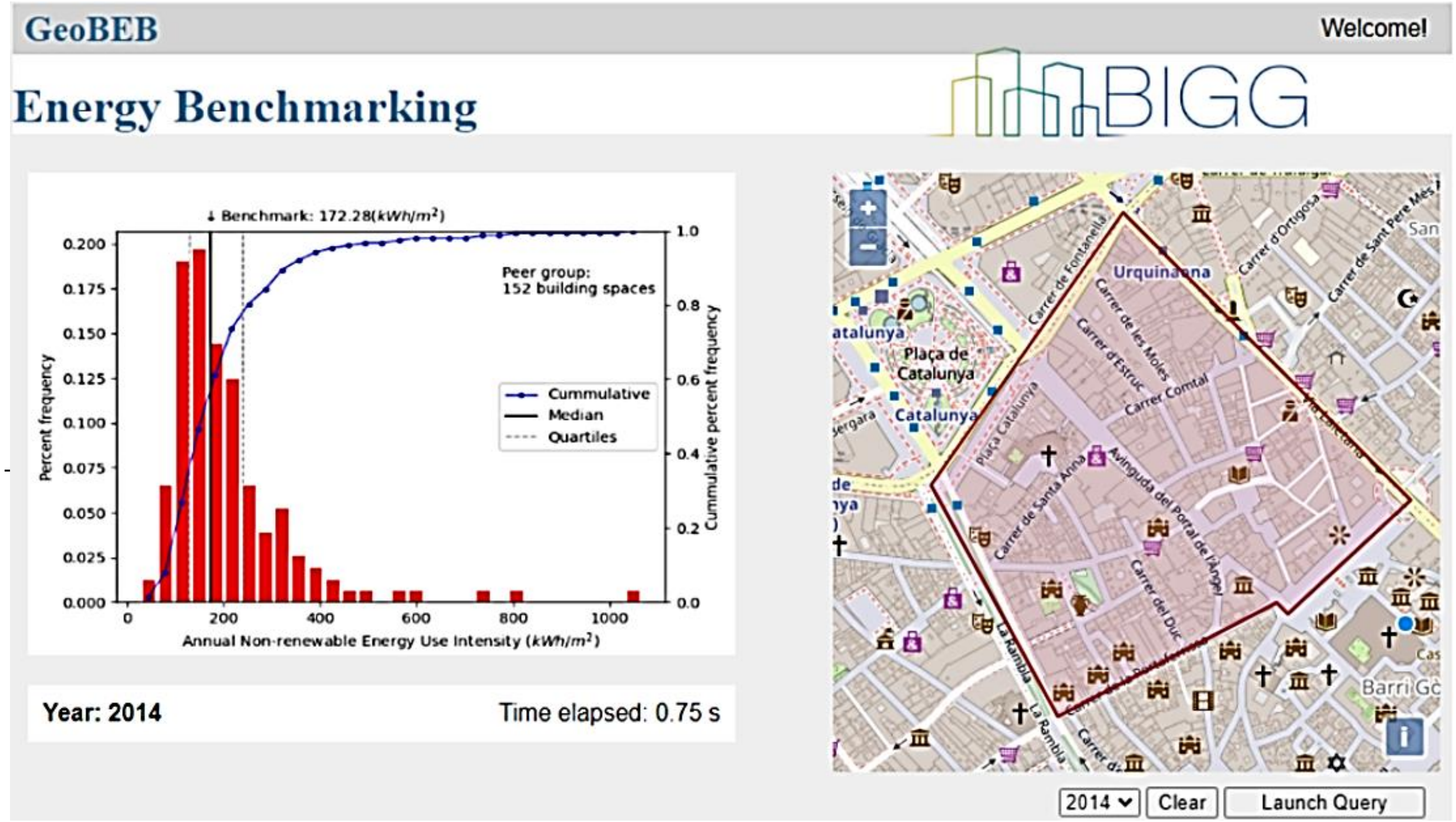


Figure 10 – Example of energy benchmarking data with OpenStreetMap

Use case 4: Adoption of the sustainability indicators of common EU framework Level(s) in building certification

This use case does not require end-user specific visualisation.

Use case 5: Interoperability between BIM, BMS, CMMS and building simulation engines

The final customer visualization of this use case has been integrated into the application developed in use cases 1 and 2. A view has been deployed in the section "individual buildings form" where the final user can explore the subzones of the buildings, the elements contained in each of the zones, and, also, the work orders associated with these elements. The following figure shows this view.

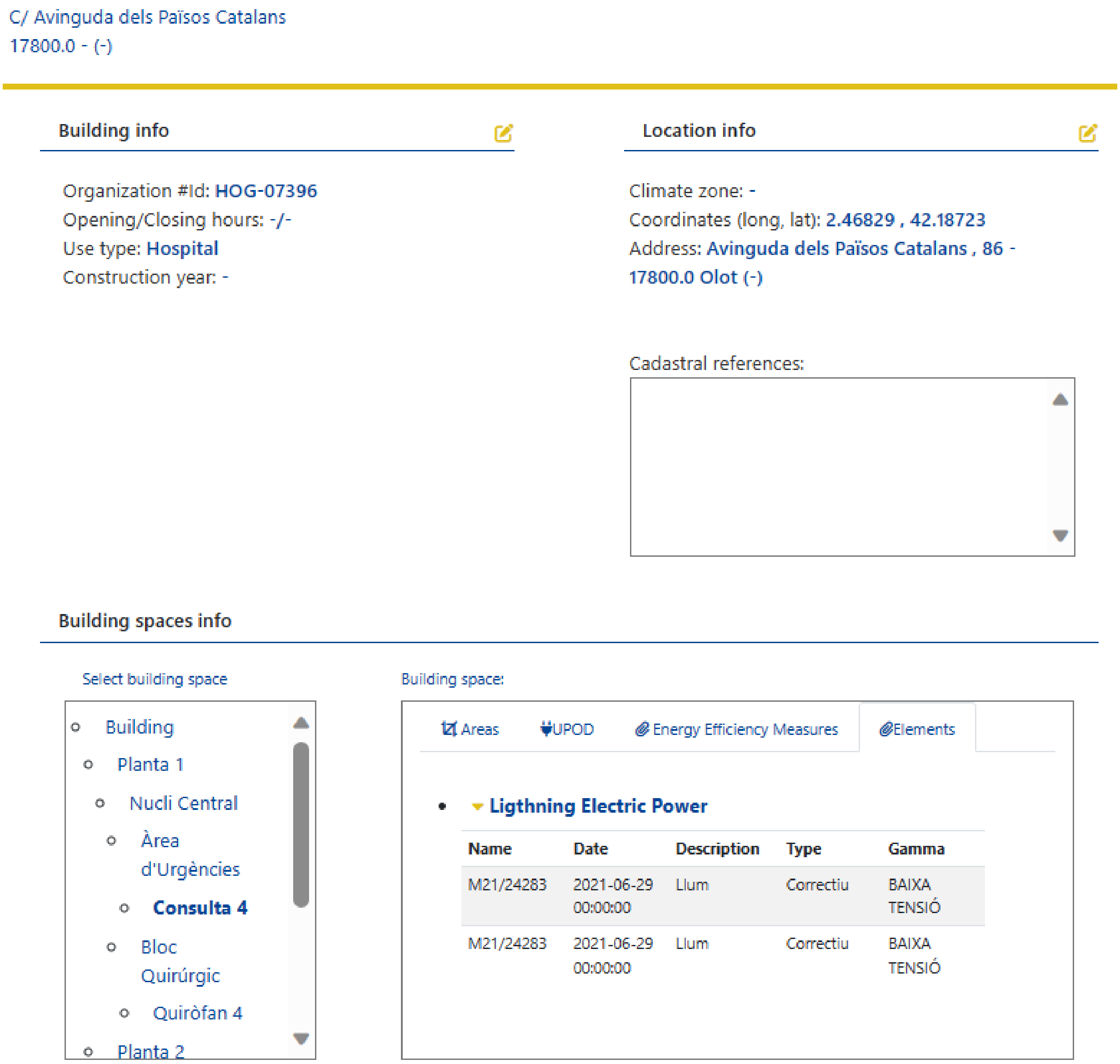
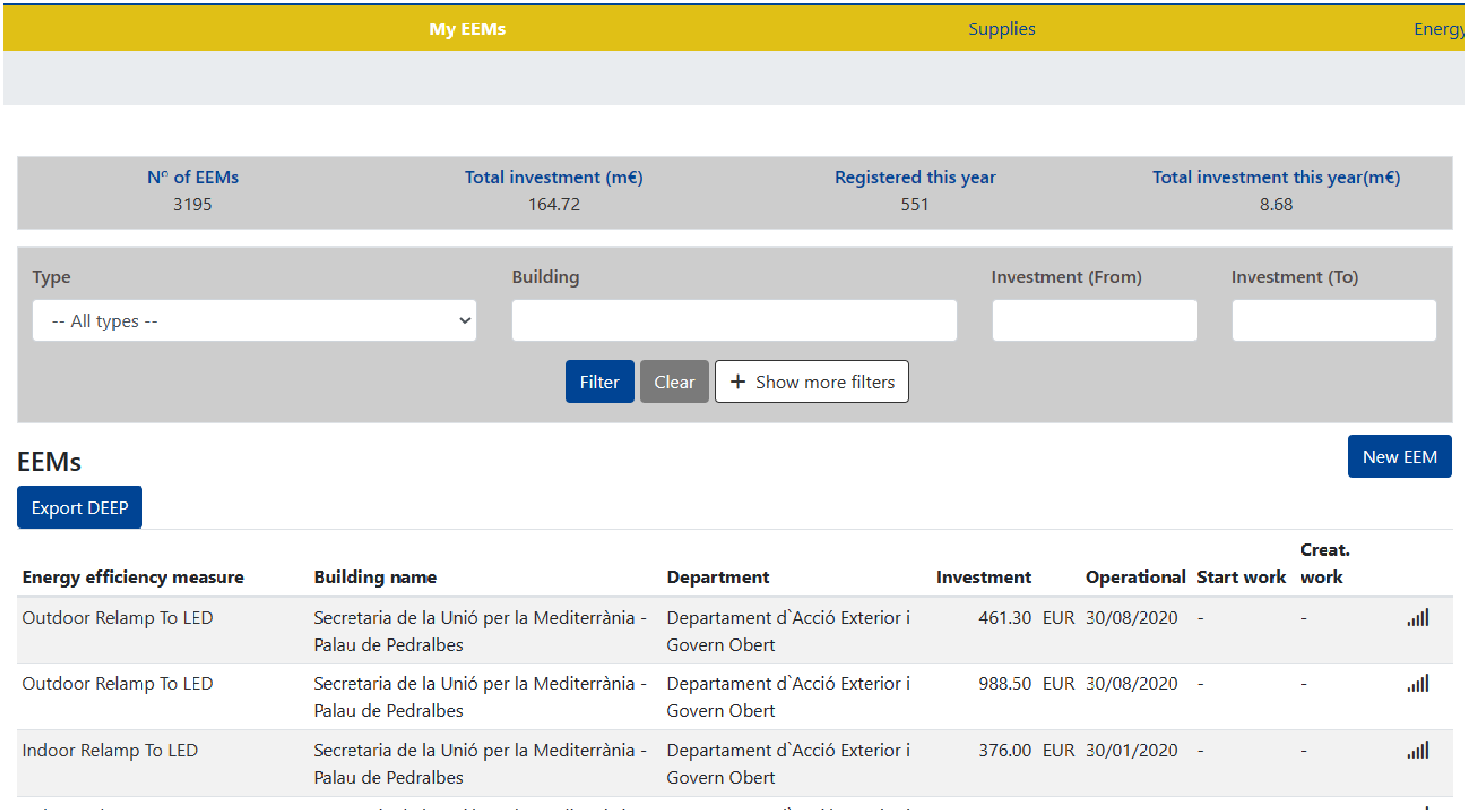


Figure 11 – Customer visualization

Use case 6: Interoperability of BIGG with EEFIG-DEEP

To ensure the integration with DEEP, CIMNE has enabled a button to export the Energy Efficiency Measures (EEMs) data in the format agreed with the DEEP platform. This button has been located in the EEMs List panel of the application.

The following image shows the export button (a) together with the file it generates (b).



a) A button to export the Energy Efficiency Measures (EEMs) data

Simple data entry

About You, the data provider	
Your name	
Your email address	
Your telephone number	
Your organisation	
What is the nature of your organisation?	

Coloring	
Essential	Drop down boxes
Confidential	Double click cells for multiple selection
Essential and Confidential	
Not in use	

Hide column comments

Show column comments

DEEP DE-RISKING ENERGY EFFICIENCY PLATFORM

Project		Where is the investment located?		---------------------- Sector Indicators ----------------------						---
(1) Project ID	(2) Project Titl	(3) Country	(5) City/locality	(12) Is the investment in a building, in industry, or in infrastructure?	(13) Industry Sector/Organisation type	(14) Organisation size	(15) Building type	(16) Ownership	(17) Floor area of building m2	
		ES	Santa Coloma de Gramenet	Building	Public administration and defence		Health care		9.550,35	ElectricPowerSystemMeasure ElectricEquipmentMeas
		ES	Lleida	Building	Public administration and defence		Public buildings		6.491,17	LightingMeasure LightingIndoorMeasure IndoorRelan
		ES	Gavà	Building	Public administration and defence		Health care		4.393,00	BuildingFabricMeasure DoorsMeasure DoorsReplace
		ES	el Prat de Llobregat	Building	Public administration and defence		Health care		-	ManagementMeasure BuildingEnergyManagementSys
		ES	Barcelona	Building	Public administration and defence		Health care		5.248,01	ElectricPowerSystemMeasure ElectricEquipmentMeas
		ES	el Prat de Llobregat	Building	Public administration and defence		Health care		-	HVACAndHotWaterMeasure HotWaterSystemMeasure
		ES	el Prat de Llobregat	Building	Public administration and defence		Health care		-	ElectricPowerSystemMeasure ElectricEquipmentMeas
		ES	Sant Llorenç Savall	Building	Public administration and defence		Health care		9.896,25	HVACAndHotWaterMeasure CombinedHeatingCooling
		ES	el Prat de Llobregat	Building	Public administration and defence		Public buildings		4.321,00	BuildingFabricMeasure WindowMeasure WindowRep
		ES	Mollerussa	Building	Public administration and defence		Health care		4.911,60	ElectricPowerSystemMeasure ElectricEquipmentMeas
		ES	Móra d'Ebre	Building	Public administration and defence		Health care		4.664,44	ElectricPowerSystemMeasure ElectricEquipmentMeas
		ES	Tortosa	Building	Public administration and defence		Office buildings		1.271,19	ManagementMeasure BuildingEnergyManagementSys
		ES	l'Hospitalet de Llobregat	Building	Public administration and defence		Industry		3.517,53	ManagementMeasure BuildingEnergyManagementSys
		ES	Tarragona	Building	Public administration and defence		Office buildings		799,89	ManagementMeasure BuildingEnergyManagementSys
		ES	Valls	Building	Public administration and defence		Office buildings		2.729,00	HVACAndHotWaterMeasure CoolingSystemMeasure C
		ES	Barcelona	Building	Public administration and defence		Office buildings		12.171,75	LightingMeasure LightingIndoorMeasure IndoorRelan
		ES	Abrera	Building	Public administration and defence				2.370,10	ManagementMeasure BuildingEnergyManagementSys
		ES	Barcelona	Building	Public administration and defence		Office buildings		19.633,95	HVACAndHotWaterMeasure CoolingSystemMeasure C
		ES	Barcelona	Building	Public administration and defence		Public buildings		927,00	HVACAndHotWaterMeasure CombinedHeatingCooling
		ES	Arenys de Mar	Building	Public administration and defence		Public buildings		-	RenewableGenerationMeasure SolarPhotovoltaicInst
		ES	Granollers	Building	Public administration and defence		Public buildings		1.076,00	LightingMeasure LightingIndoorMeasure IndoorRelan
		ES	el Prat de Llobregat	Building	Public administration and defence		Public buildings		508,00	HVACAndHotWaterMeasure CombinedHeatingCooling
		ES	Blanes	Building	Public administration and defence		Health care		1.156,25	HVACAndHotWaterMeasure CombinedHeatingCooling
		ES	Sant Cugat del Vallès	Building	Public administration and defence		Office buildings		28.774,70	HVACAndHotWaterMeasure CoolingSystemMeasure C
		ES	Tàrrega	Building	Public administration and defence		Public buildings		1.636,00	HVACAndHotWaterMeasure CoolingSystemMeasure C
		ES	Sant Adrià de Besòs	Building	Public administration and defence		Educational buildings		1.897,00	HVACAndHotWaterMeasure HeatingSystemMeasure H

b) The file that is generated from the export button

Use case 7: Interoperability between EU Building Stock Observatory (EUBSO) and national/regional Energy Performance Certification (EPC)

This use case does not require end-user specific visualisation. The integration has been done directly into the database by linking the EPCs data with the EUBSO data.

HELEXIA User Interface

For Business Cases 4 and 5 (BC4 & BC5), Cordia and Helexia relied on two different user interfaces: Energis.Cloud and Grafana.

III.1.2.a.1. Energis.Cloud

The Energis.Cloud platform was used as the main platform to build the BC4 solution as a dedicated set of dashboards (interactive screens displaying the relevant data and KPIs) and reports (i.e. automatic generation of PDF reports). It was also used to store time series data (typically data about consumption, weather, indoor comfort, occupancy) and models but also to compute KPIs such as energy savings from EPC projects, bonus and malus fees, etc.

In details, the Energis.Cloud platform offers the following functionalities that were key to the success of BC4 & BC5:

- Integration of all the physical asset related information under the form of an interactive asset tree (see Figure 122).

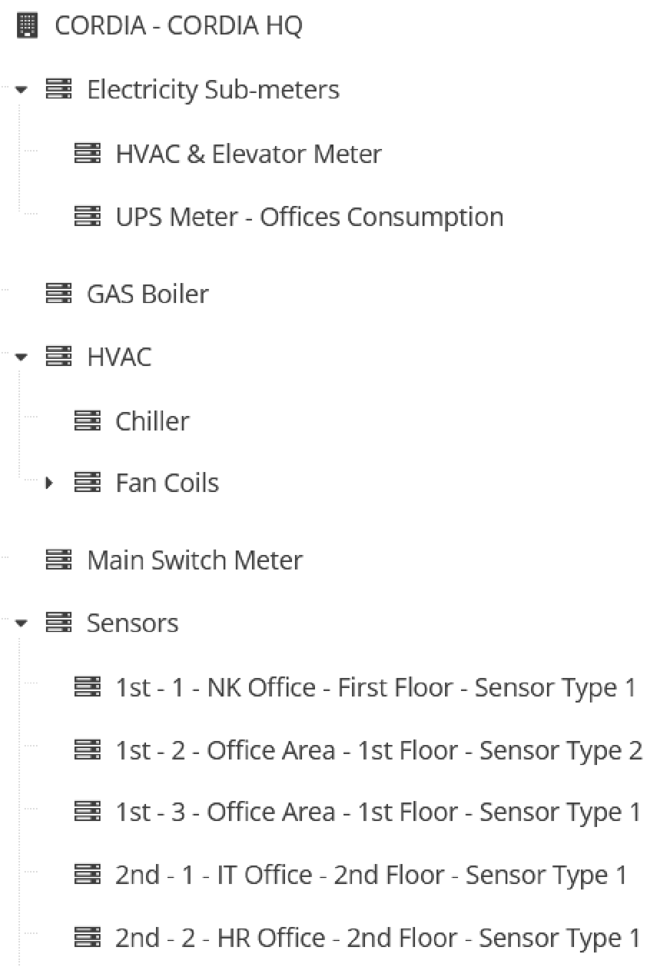


Figure 12 – An example of building structure in Energis.Cloud

- Association of all asset and contract related information in a detailed and flexible way.
- Collection of all required time-series data such as consumption, weather, sensors or actuators data, also including semi-static data such as prices or time-varying distribution keys.
- Flexible calculation of KPIs based on this data such as degree days, generated savings, comfort ratios, bonus/malus savings distributions, etc.

- Generation of regression data models using an ad hoc tool (non-generic) pre-existing in the platform (Figures 13) or integrating the model identification tools created in WP5 from the BIGG AI Toolbox[17]. Additionally, the platform allows to integrate custom made regression models.

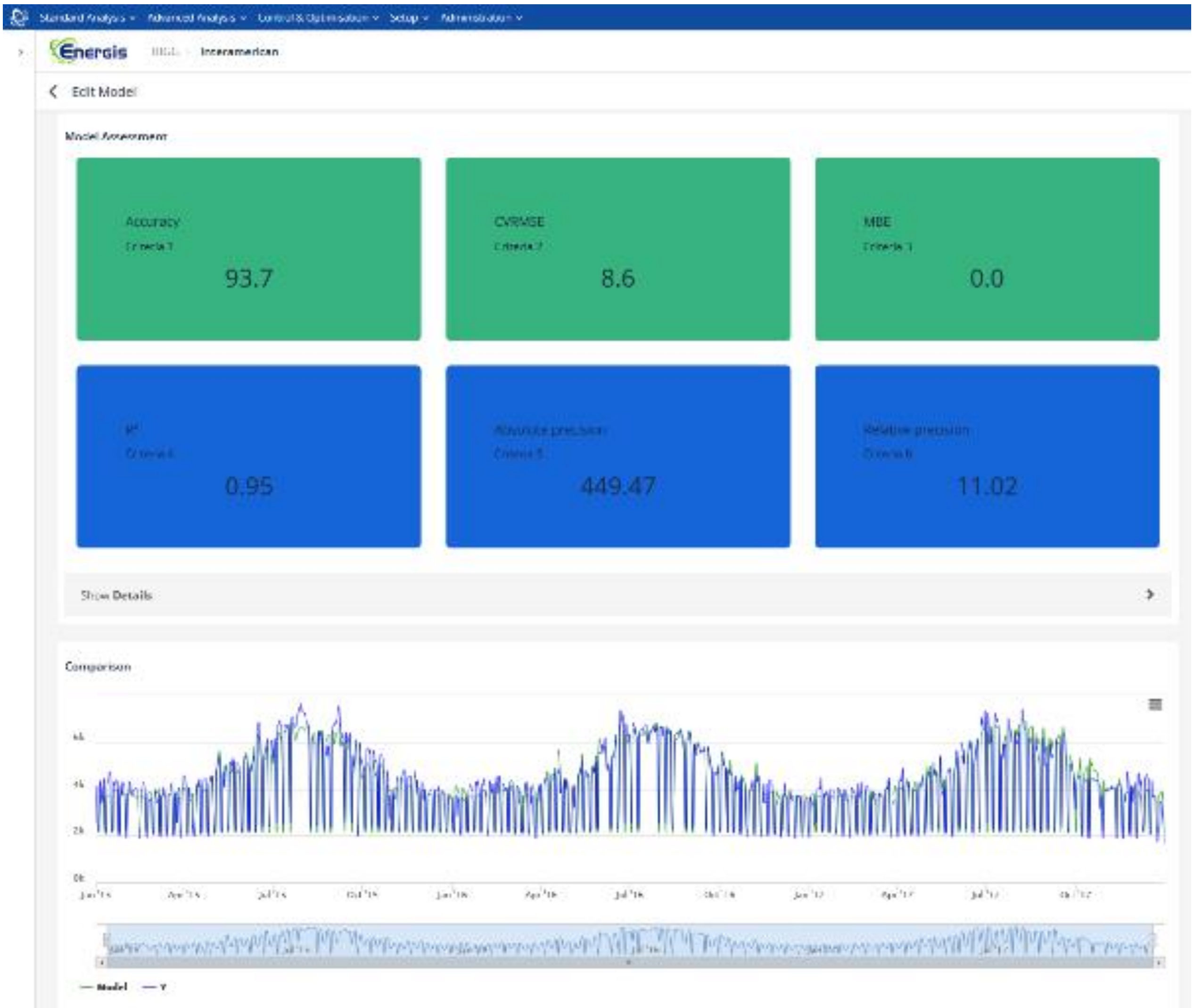


Figure 13 – Energis.Cloud has a built-in regression modelling engine able to identify models for a site in an ad hoc way.

- Construction of generic dashboard templates, built in a way that the same dashboard can be used for any BC4 or BC5 project with no additional work.
- Automatic generation and sending of pdf reports, following the same logic as for the dashboards, that is, requiring no extra work when new BC4 or BC5 projects are integrated.

Below are some examples of dashboards that have been created for BC4 and BC5 using the Energis.Cloud platform.

- The savings follow up screen where the savings of an Energy Performance Contract (EnPC) savings can be tracked compared to a target and bonus/malus situations can be identified (see Figure 14).

[17] More details about the BIGG AI Toolbox for buildings can be found in the deliverable D5.2.

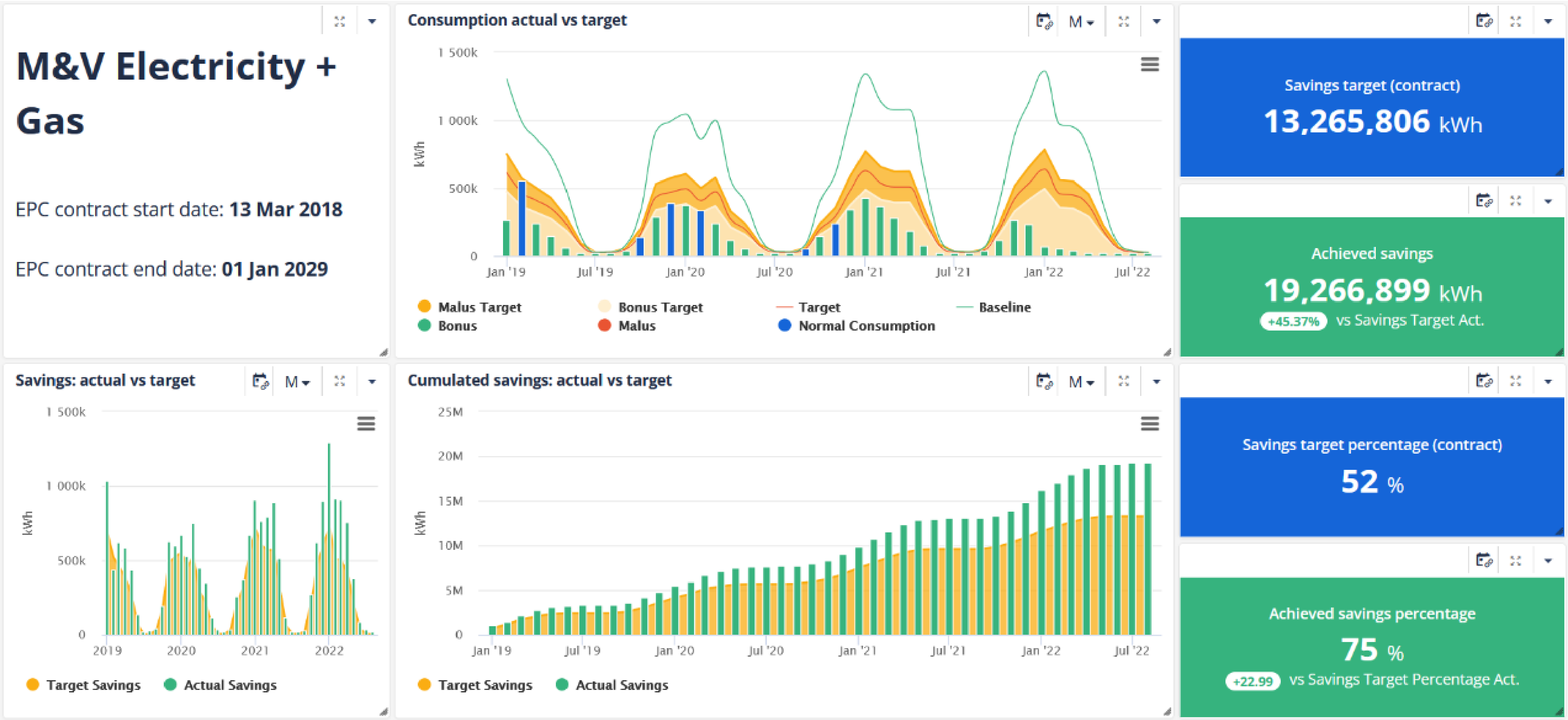


Figure 14 – Energy Performance Contract to track savings.

On the top chart of Figure 14, the actual consumption of the site (histograms) is compared to the baseline – historical data - (the green line) and to the EnPC' project target (the red line). Depending on the contract logic, there can be a bonus/malus mechanism based on the achieved performance. This is represented in the chart as well: within the orange area, the performance is considered to be as expected (the actual consumption is shown in blue), below it, over-performance is achieved (the actual consumption is shown in green), while above it, under-performance is observed (the actual consumption is shown in red, not in the illustrated case though).

- The Financial gains of an EnPC from the point of view of the Energy Services Company (ESCO) financing the project (see Figure 15, left part).
- The Financial gains of an EnPC from the point of view of the end customer (see Figure 15, right part).
- The comparison of regression models from different origins, typically custom made (with Excel), identified from the Energis.Cloud ad hoc tool or identified from the BIGG AI Toolbox modelling tool (see Figure 16).

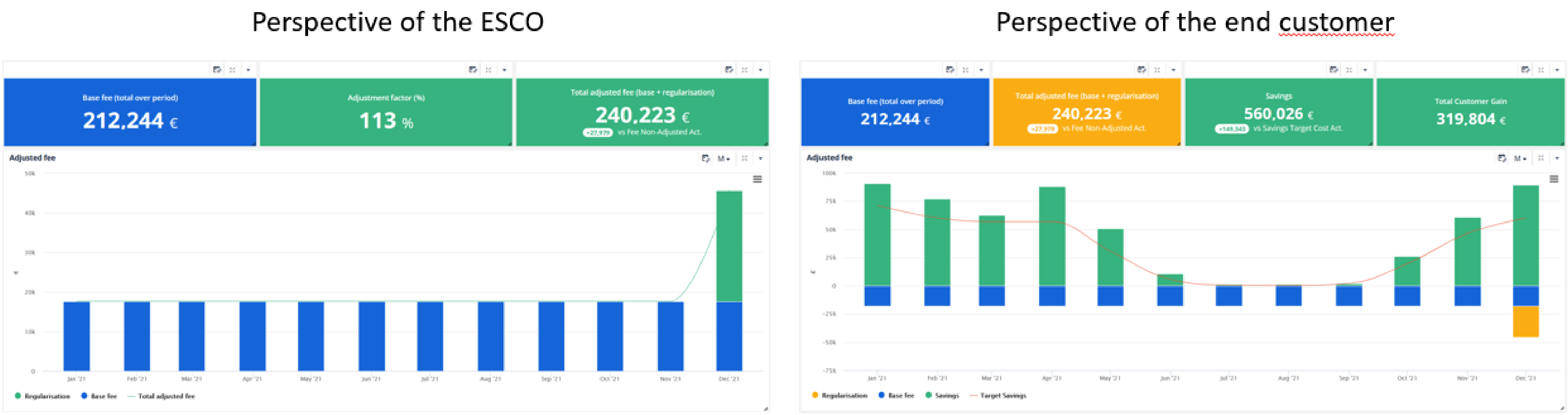


Figure 15 – Financial gain tracking from ESCO and customer side.

On the left panel of Figure 15, the view on what the ESCO will perceive as benefits from the project, including a bonus in this case at the end of the 12-month period. On the right, the perspective of the end customer with its costs and its benefits originating from the achieved consumption savings.

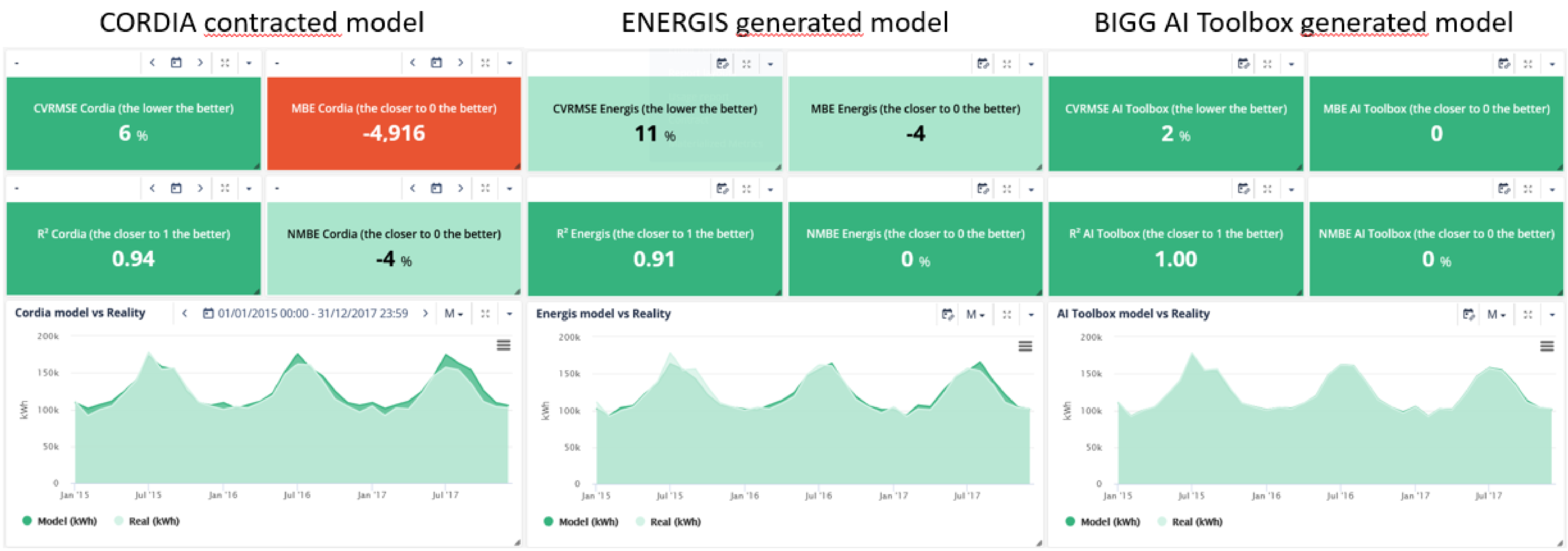


Figure 16 – Comparison analysis of CORDIA, ENERGIS and AITB models.

In Figure 16 we show the comparison between CORDIA, ENERGIS and AITB models. The model generated from the BIGG AI Toolbox provided significantly more accurate results than the other two models. Visually, the model curve in all the three cases is practically superposed to the actual data. So, one can visually compare where the overlap is the best.

- The follow up of comfort conditions in a building where control actions are taken (see Figure 17).

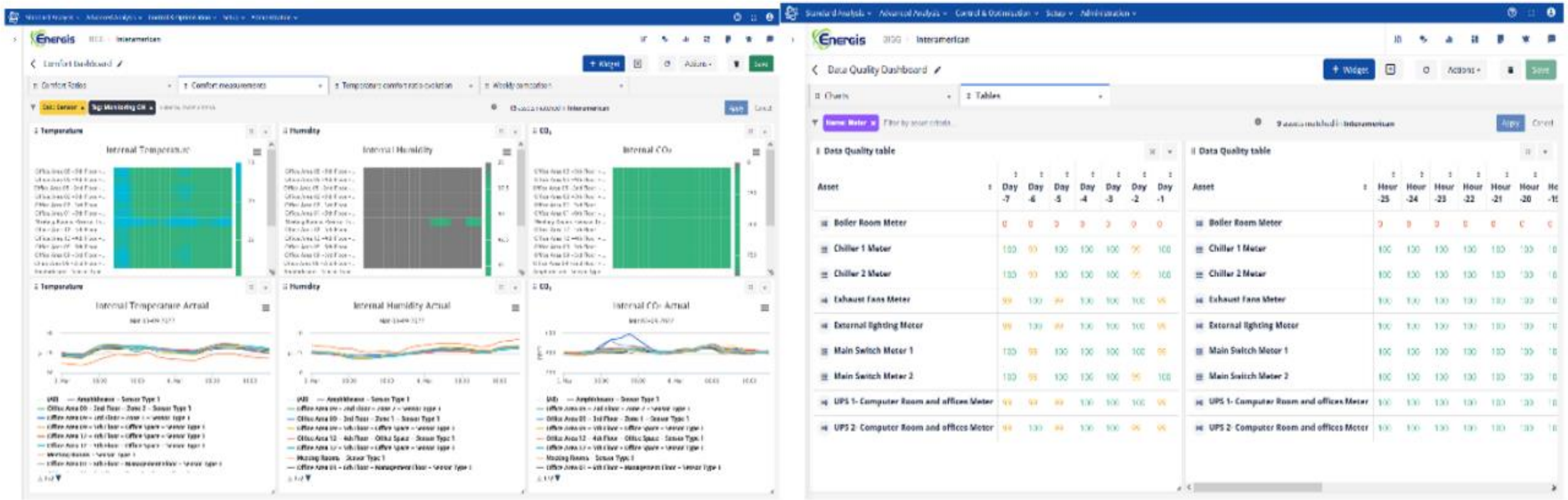


Figure 17 – Comfort conditions.

- Comfort conditions are analysed in detail for every sensor and at a global scale using comfort ratios that are aggregated across all sensors and taking the occupied hours and comfort thresholds into account
- The general monitoring of meters and sensors.
- The reporting of energy savings as an automatically generated monthly PDF report (see Figure 18).

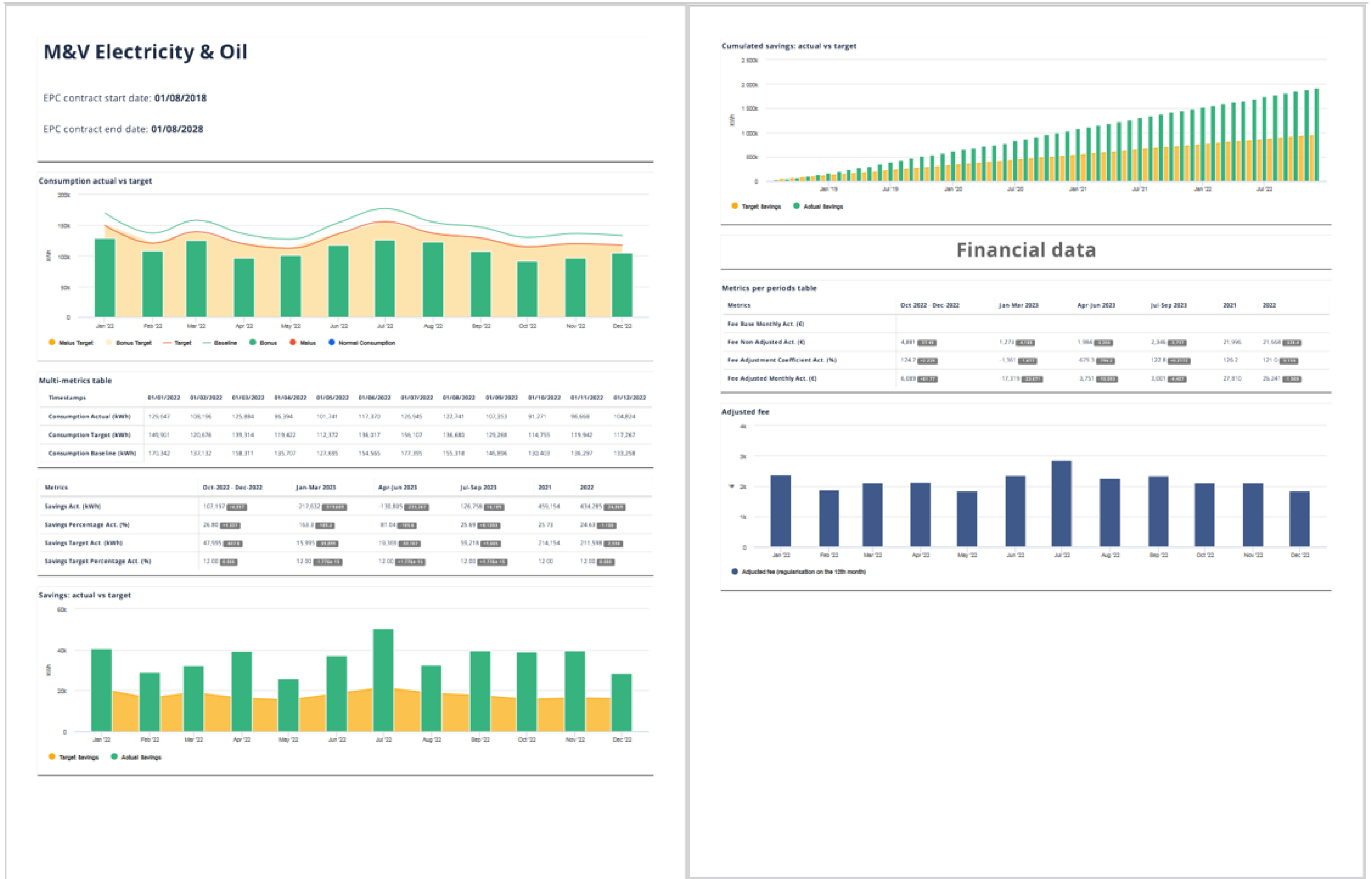


Figure 18 – A typical generic report generated by the solution to report the savings of the latest period.

The Energis.Cloud platform is developed by ENERGIS and was already a working commercial platform before the BIGG project. However, to comply with the objectives of the BIGG project, several improvements of the platform had to be developed among which:

- The integration of the AI Toolbox generated models which, unlike the built-in Energis.Cloud modelling tool, is able to identify models for a large set of sites, taking their specificities automatically into account. This integration required several changes at the core of Energis.Cloud:
  - Review the way Energis.Cloud metrics are working, allowing to integrate external regression models as formulas and extending the APIs to create and define these metrics.
  - Standardize the asset properties, in agreement with the BIGG ontology, ensuring that the AI Toolbox will always find the input information it needs such as weather, occupancy or consumption data but also some building characteristics. As mentioned in Section II.2.2. , this was achieved by introducing new “asset roles” and “meta-metrics” concepts[18].
  - Integration of the weather data from the Weatherbit service as a microservice, Cfr. Section II.2.2.
- The standardization of asset properties through asset roles and meta-metrics was also required for general harmonization of the project’s data with respect to the BIGG ontology.
- Several dashboard improvements were introduced including a new, more flexible date picker which was required to define the necessary “contract length” periods needed for the BC4 solution. Alongside this change, a new zooming mechanism, synced with the new periods, was introduced as well as several widget improvements, required for retro-compatibility.
- A new pre-computation mechanism called “metric materialisation” has been introduced to cope with the large volumes of data that are expected to be encountered once the developed solution are pushed to larger scales.

III.1.2.a.2. Grafana

Grafana, a second User Interface, was used in the context of BC5. The Grafana framework presents a significant flexibility in terms of features and is used to display real time data where the Energis.Cloud dashboard mainly focuses on displaying relevant KPIs. The Grafana dashboard displays the HVAC equipment operating conditions (boiler status, water loop temperature, chiller status…) and control status both used as inputs by the rule based engine.

[18] The Asset roles and Meta-metrics concepts have been later relabeled into Metric groups and Metric definitions, with an additional concept of Metric templates gluing the two concepts. However, for past reference purpose, we included the original names in the text. More documentation about these concepts can be provided on request.

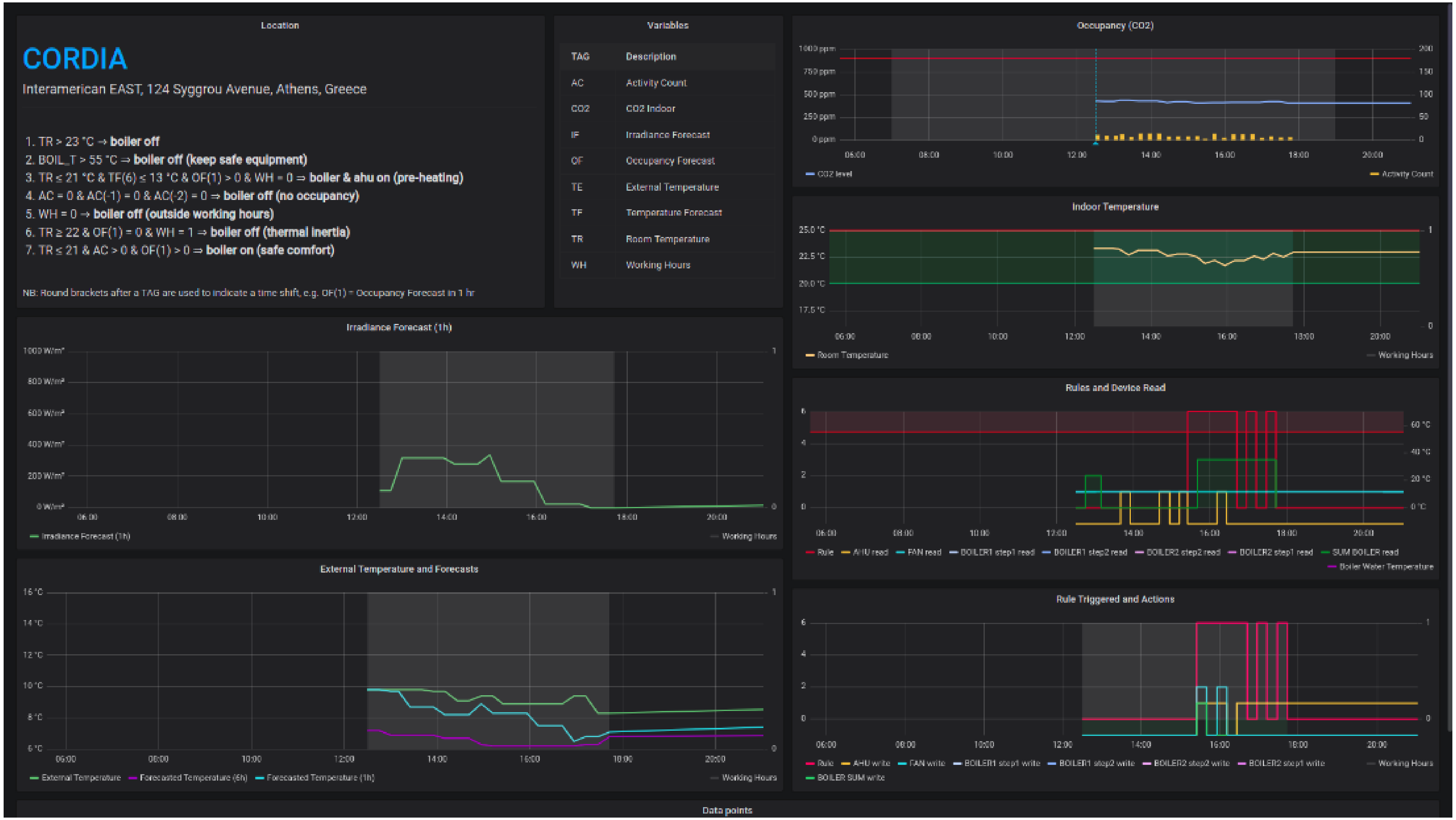


Figure 19 – Dashboard for real time tracking of the impact of rules

The dashboard in Figure 19 also displays in real time the status of rules that are implemented to allow a quick interpretation of the impact of rules implemented over the site operating conditions.

HERON User Interfaces

This use case provides the processes required to activate user flexibility for the Demand Response solution. The overall process is being monitored by HERON as the electricity supplier, parsing suggestions to its consumers on shifting their consumption to match system or market conditions.

<u>End user application</u>

Communication with end users takes place through an app. The app is designed to focus on consumption monitoring, real-time and historical data and to integrate daily advice implementing the Demand Response service. Currently, users have access to a Grafana dashboard which provides data on electricity consumption per phase and aggregated, for energy (kWh), power (KW), voltage (V) and Power Factor.

Figure 20 – Historical data for Power, Voltage, Energy and Power Factor per phase

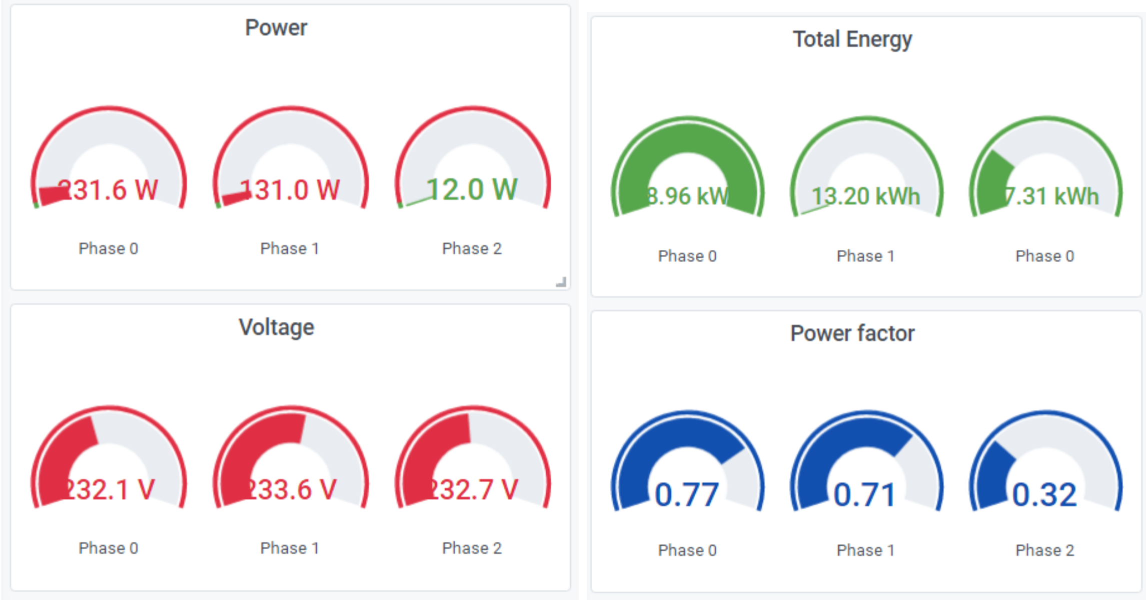


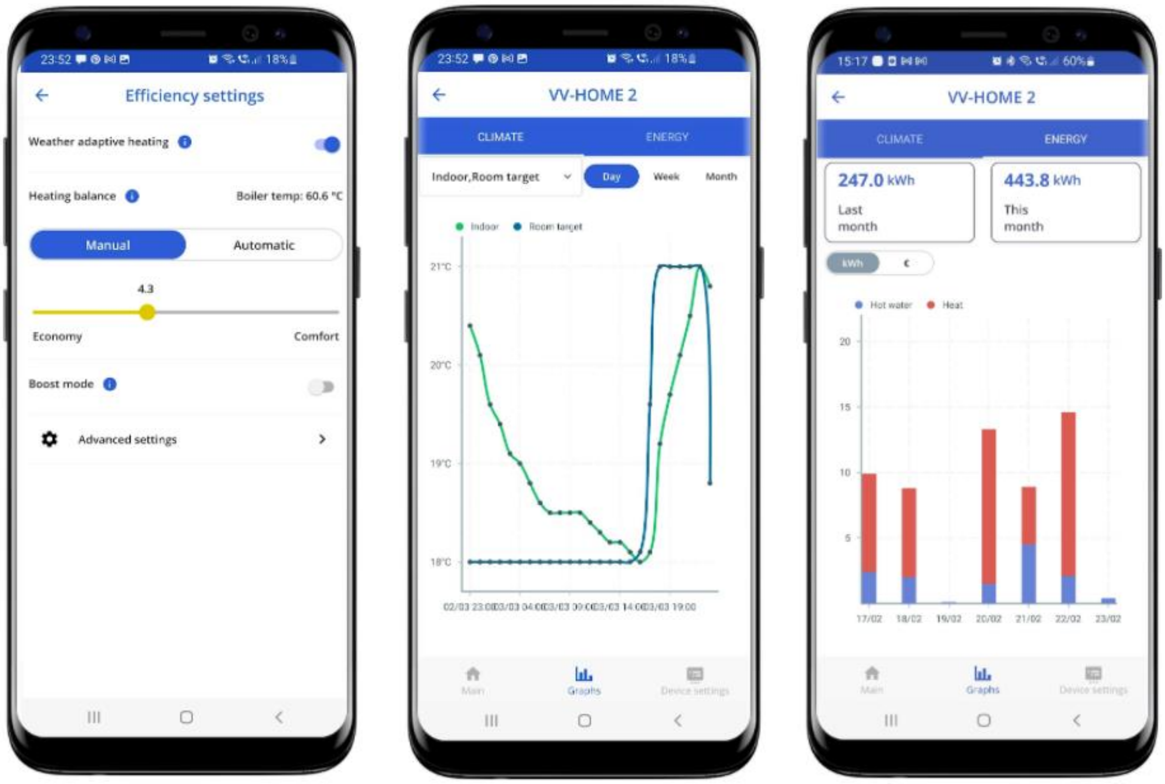
Figure 21 – Real-time data for Power, Voltage, Energy and Power Factor

DomX User Interfaces

This use case provides all the processes required for enabling Gas Suppliers to deliver Energy Efficiency and Flexibility Management services for legacy space heating gas boilers of residential consumers. The overall process progress is being monitored and managed both by the supplier at the portfolio level and the participating consumers at the individual household level.

<u>End user Application</u>

The domX smartphone application is publicly available to all end users of the domX smart heating controller product, distributed through the Google Play and App Stores. Within BIGG, the end user application is being extended to serve the needs of UC 15, by enabling end consumers to participate in gas flexibility services, while minimizing the impact of the achieved climate comfort. To this aim the app has been extended, as shown in Figure **22**, to enable end users to: a) provide their comfort limits (acceptable room temperature deviation), b) get notified about available DR events, c) opt in/ opt out to DR events, d) get informed about the achieved benefits/ comfort. Indicative screenshots from the smartphone application under development are illustrated below:



(a) (b) (c)

Figure 22 – DomX ser mobile Interfaces.

<u>Energy Supplier Dashboard</u>

The core aim of the Natural Gas supplier is to correct the daily imbalances derived from non-optimal estimation of their consumer portfolio daily demand. The overall aim of UC15 is to enable the gas supplier to dynamically adapt the demand of connected gas consumers, while offering incentives for the end consumers to participate in flexibility services. The overall process considers a wide range of available data from various sources, including building characteristics, weather forecasts, actual natural gas consumption, natural gas balancing prices and others. It can also receive data from different type of devices included in the same building, such as Energy Meters and IAQ Sensors that might impact the usage of the Natural Gas in the building (Heating from other sources, Humiditiy in the household etc.)

In the following Figure, several screenshots are illustrated to visualize the user interface developed that implements various of the functionalities presented below:

- Real-time monitoring of portfolio demand and natural gas balancing prices (Fig. 23)
- Visualization of weather forecast (Fig. 24)
- Real-time monitoring of devices (Fig. 25)
- Real-time monitoring of active boilers in map representation (Fig. 26)
- Monitoring of reported savings of the portfolio (Fig. 27)
- Remote management of heating parameters (Fig. 28)
- Monitoring for an extended set of devices for a building (energy meter, IAQ sensor, etc.) (Fig. 29)

Figure 23 – Real-time monitoring of portfolio demand

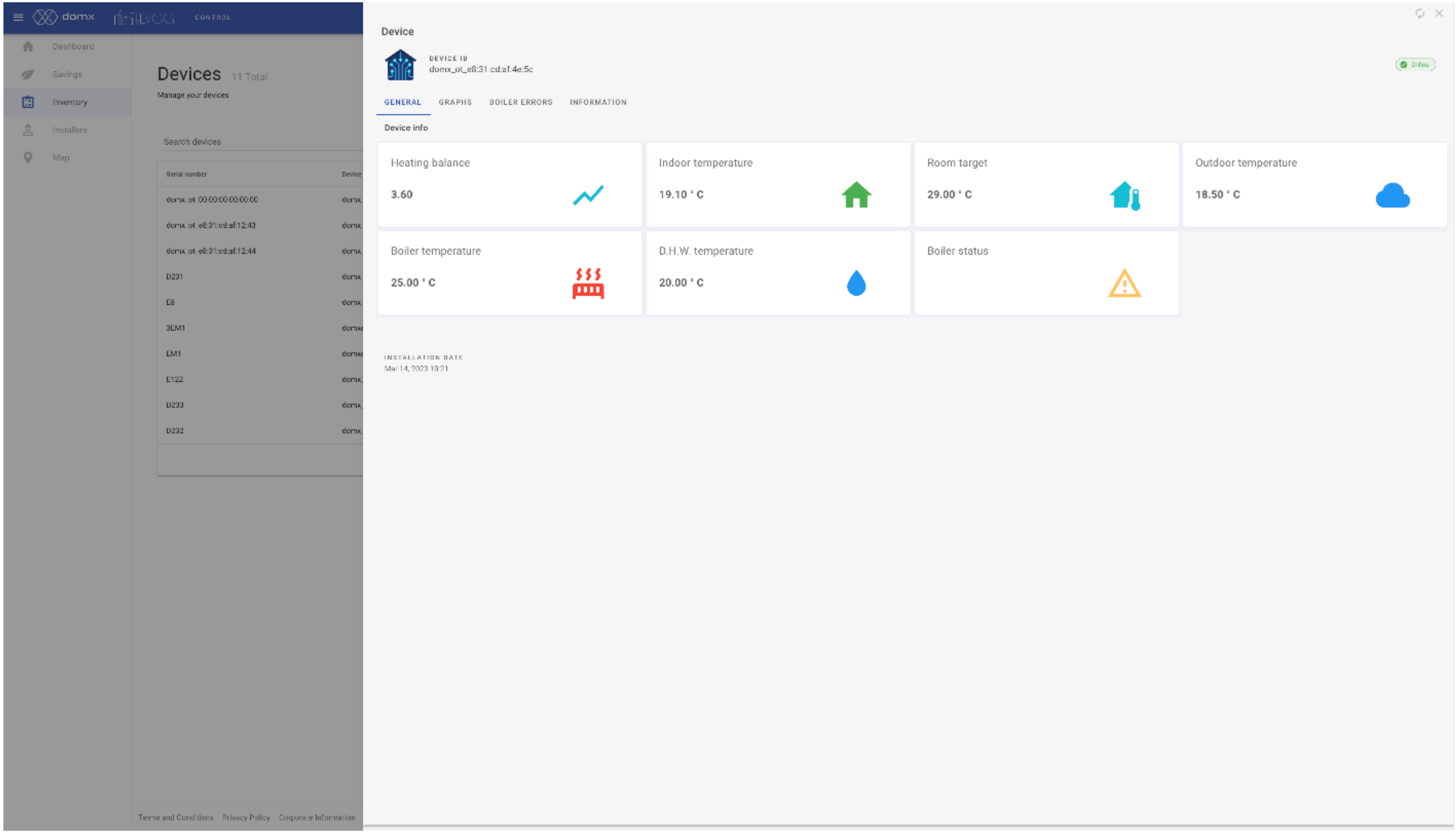


Figure 24 – Real-time monitoring of devices

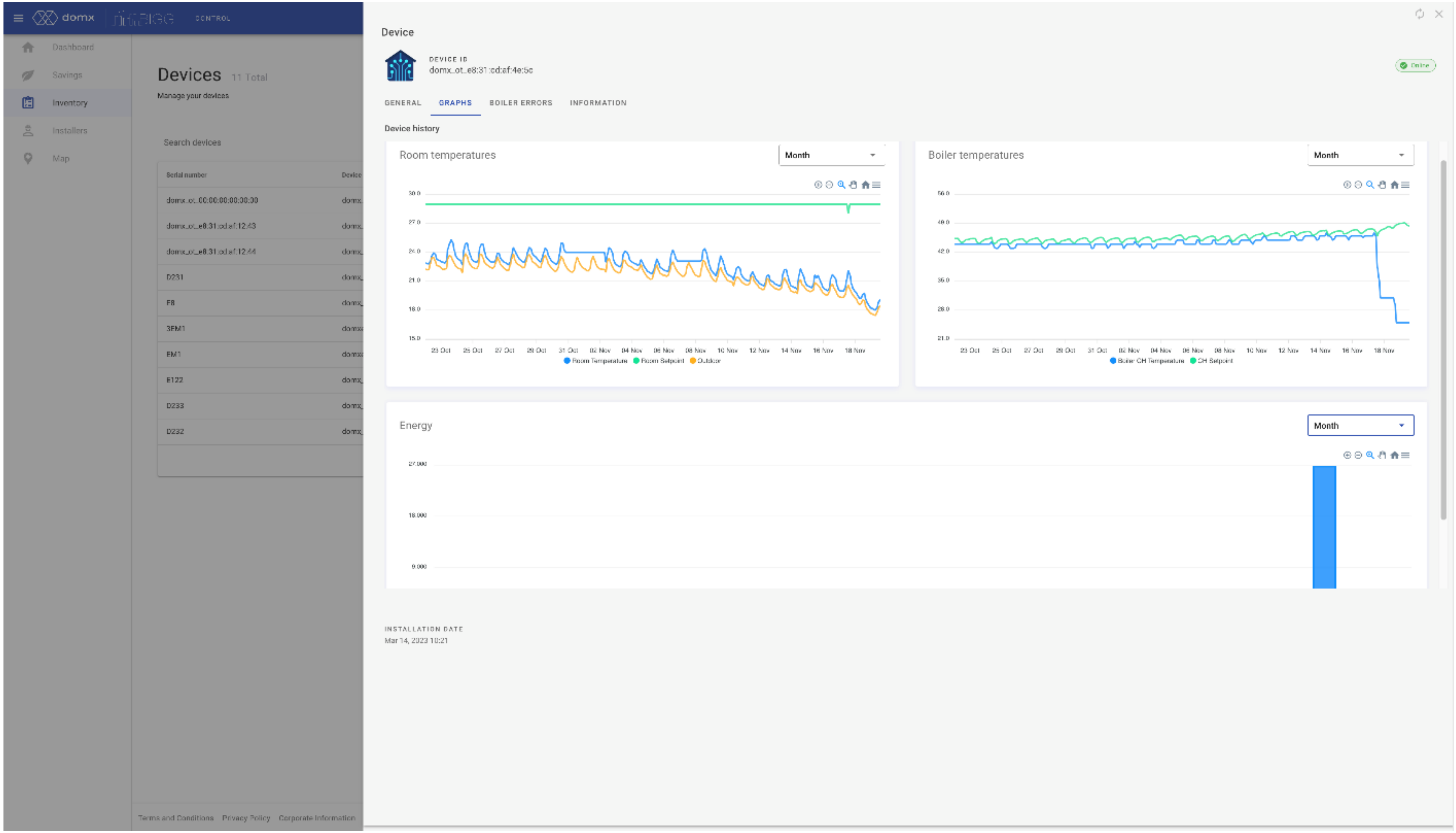


Figure 25 – Real-time monitoring of devices

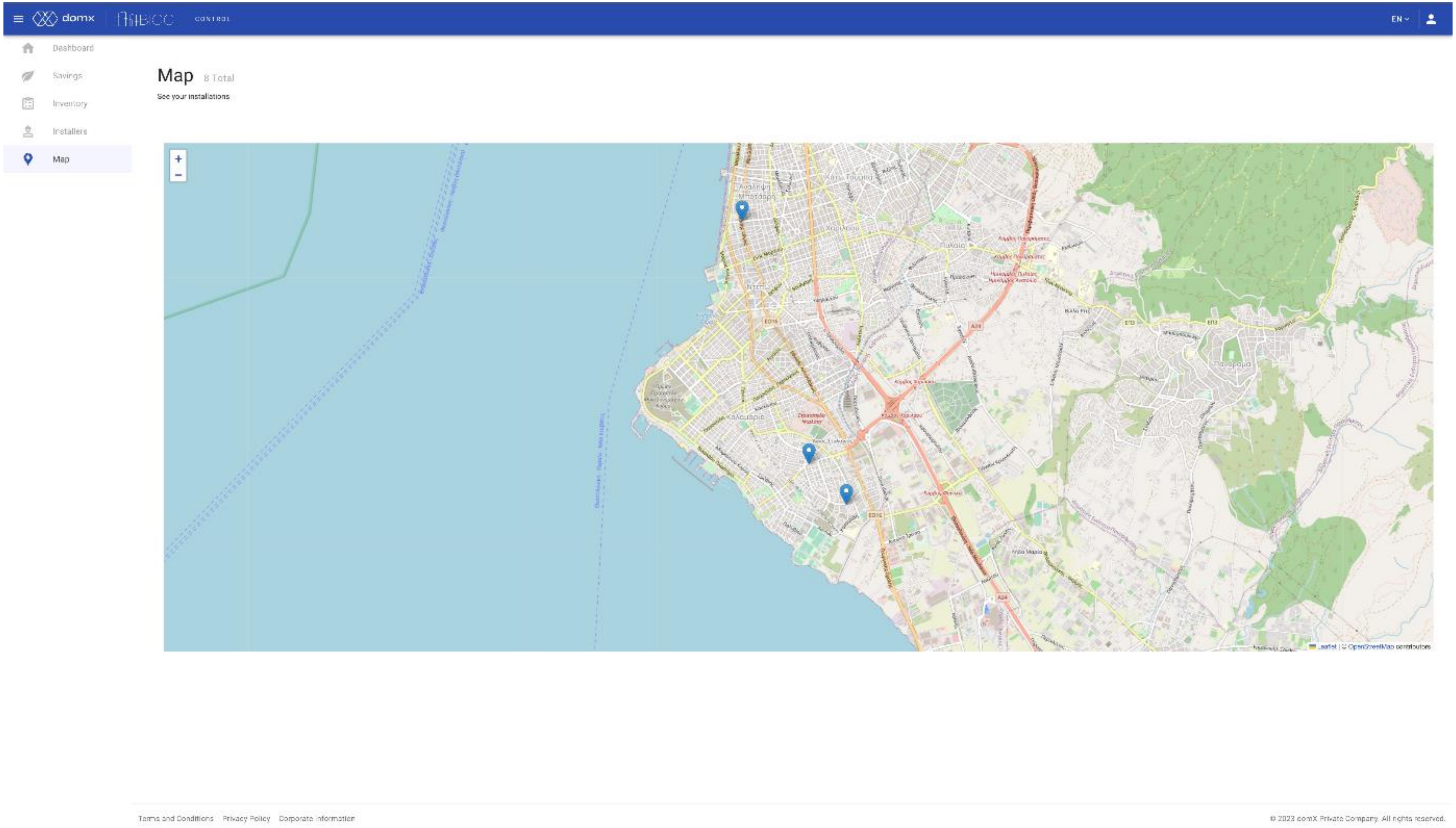


Figure 26 – Real-time monitoring of active boilers in map representation

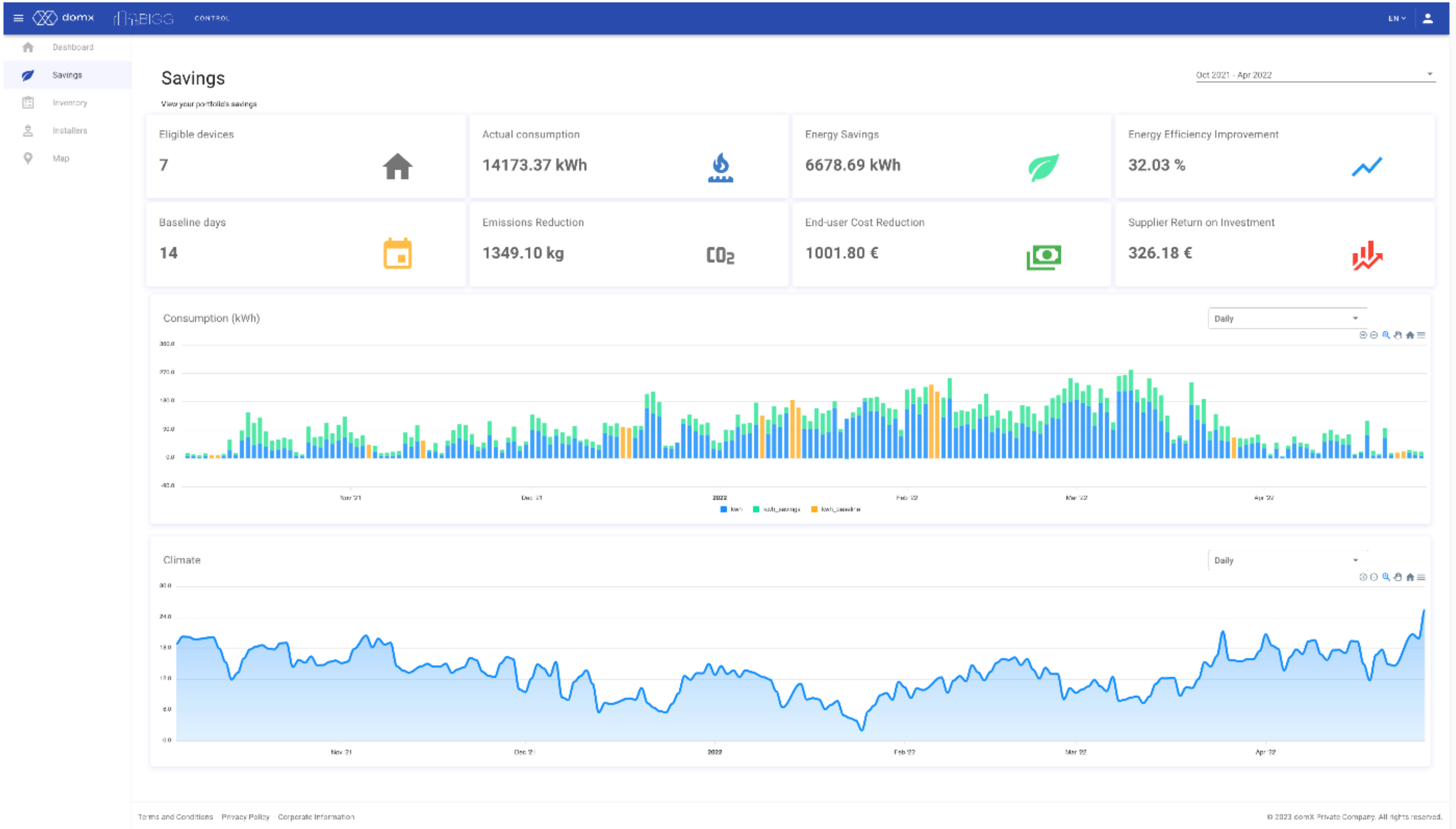


Figure 27 – Monitoring of reported savings of the portfolio (user & supplier)

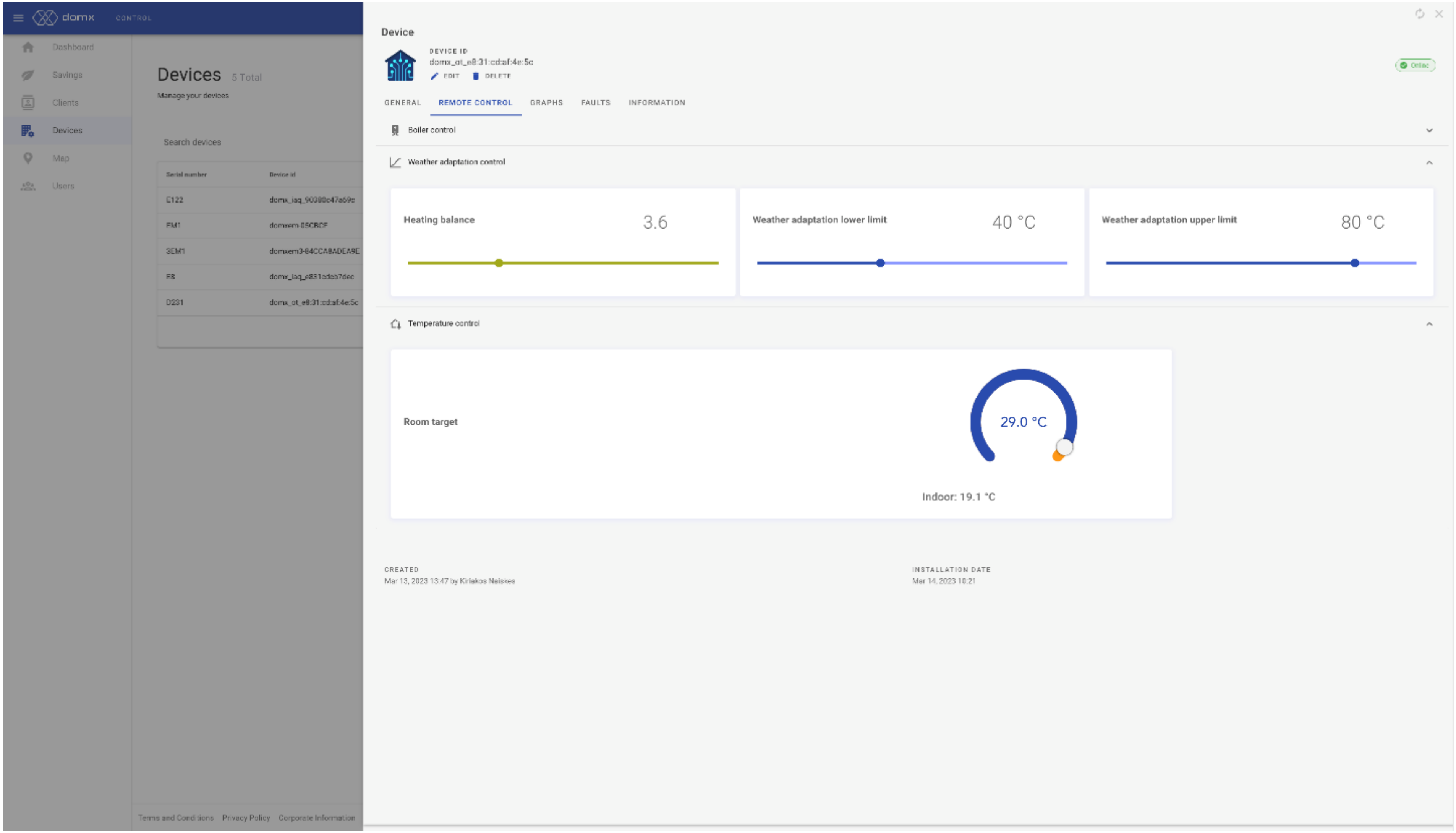


Figure 28 – Remote management of heating parameters

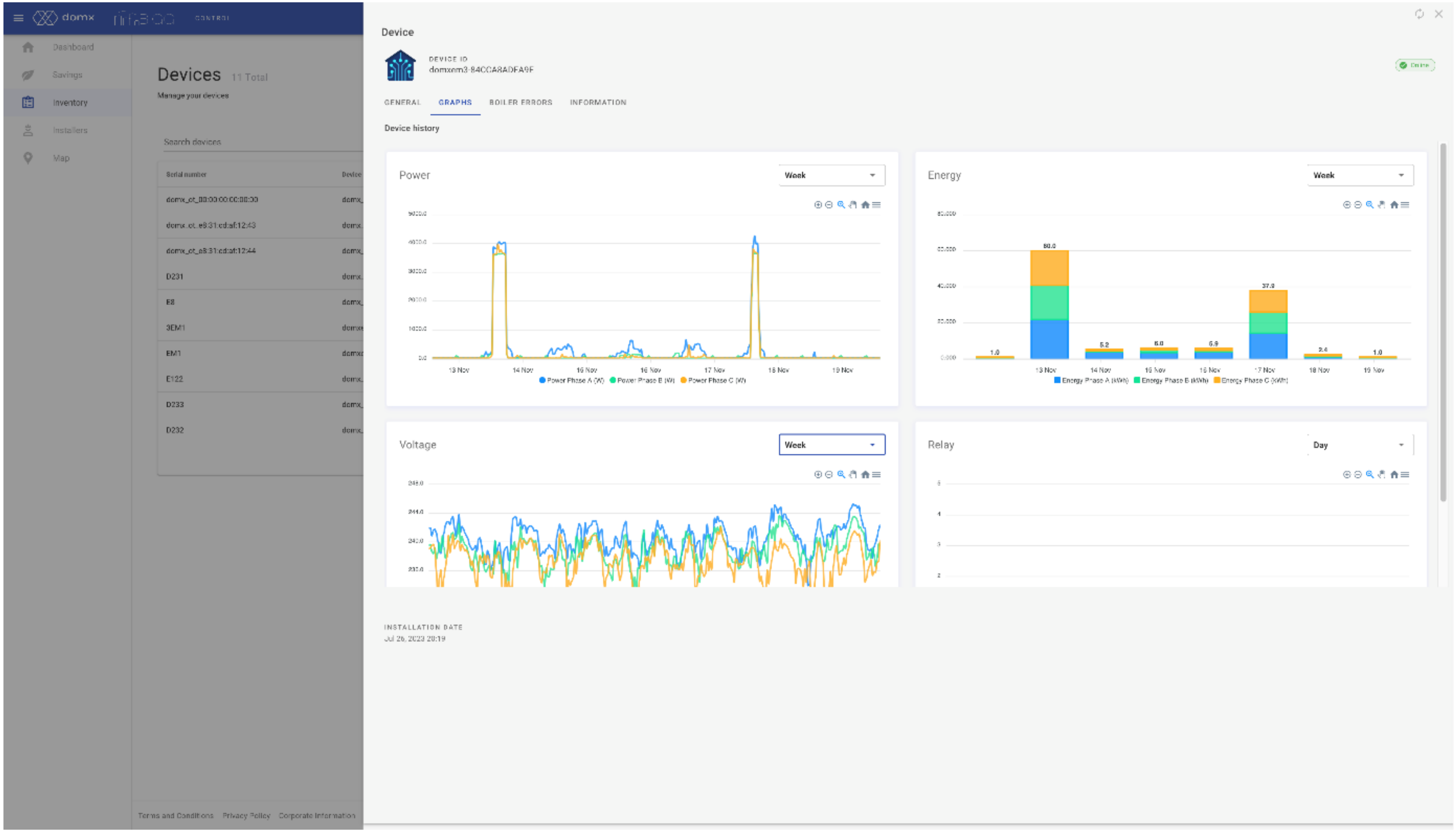


Figure 29 – Monitoring for an extended set of devices for a building (energy meter, IAQ sensor, etc.)

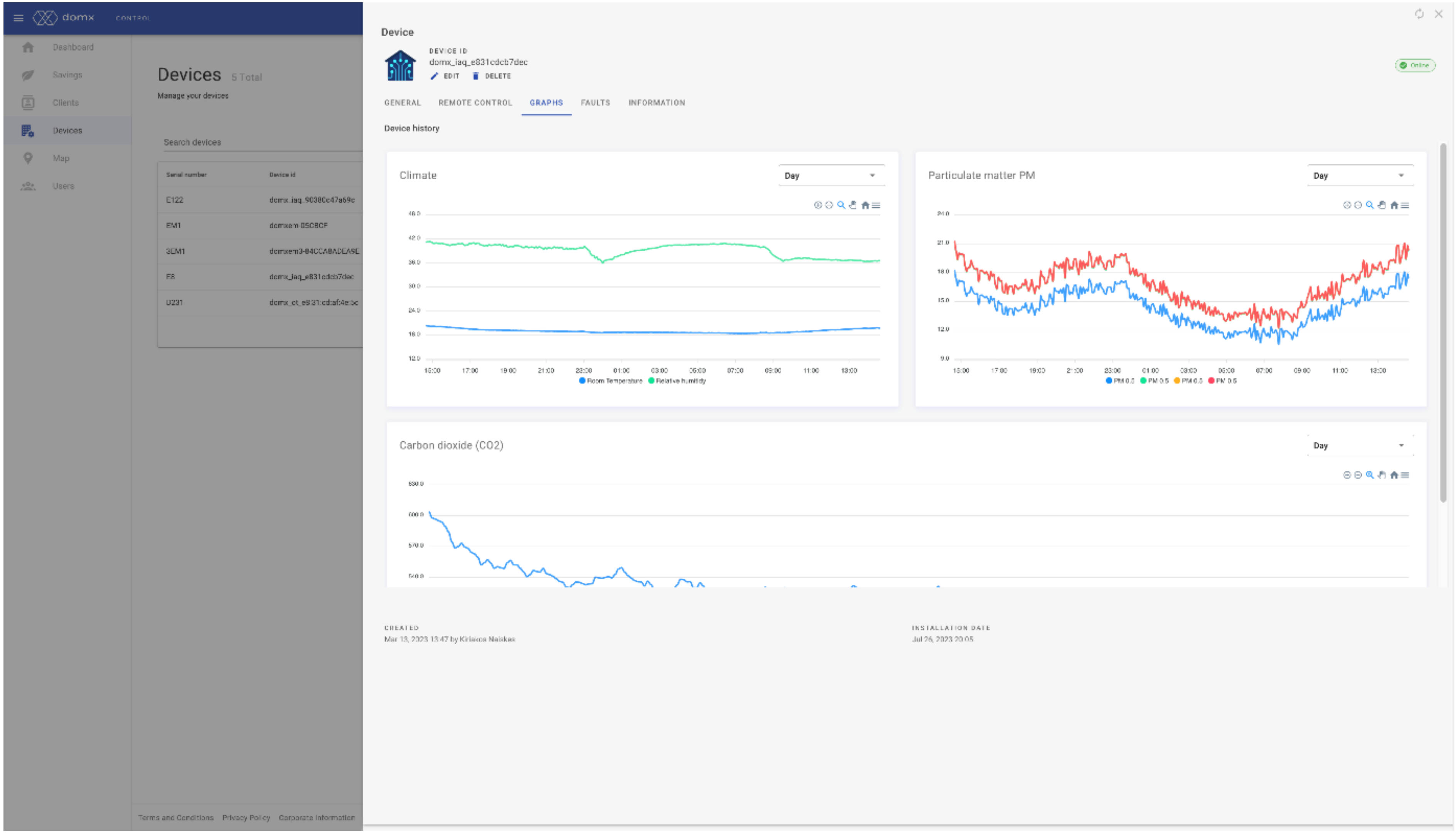


Figure 29 – Continued

IV. REPLICATIONS OF THE USE CASES

In order to facilitate the readers of this document to a comprehensive understanding of the technical setup of the use cases, we have dedicated a section to the on the replication of the pipeline setup for the actual underlying solution that connects the data to end-user dashboards. We have highlighted the material of this section in the Figure 1 of this document via a red dashed boundary. This section serves as an additional component in giving the readers a full overview without the need to consult any other deliverables to provide what functionality lies in between the communication layer, and graphical user interface embedded within our framework. Replication provides a unique vantage point, allowing readers and any other stakeholders (internal or external) to visualize and comprehend the holistic architecture of our system. By offering insights into the intricacies of the technical infrastructure, we aim to foster a deeper appreciation for the nuanced interplay between the communication layer, the user interface, and the robust security measures in place. While not a mandatory aspect for end-users, this section proves invaluable for those seeking a profound and comprehensive understanding of the underlying technology, ensuring a more informed and engaged audience.

We have provided information mostly from a technical level as providing the functional requirements are out of scope for this document. However, we highlight the important aspects of data and end product for each use case view point.

IV.1. APIs for end-users

BC6 UC 14: Electricity Demand Response

Use case 14 concerns the electricity demand response for households in Greece. This use case had the objective to forecast the energy consumption values. For UC 14, we have created two sets of API via which the end products of the pipelibe can be seemlessly integradted in the any other frameworl/architecture.

An API, or Application Programming Interface, is a set of rules, protocols, and tools that allows different software applications to communicate and interact with each other. It defines the methods and data formats that applications can use to request and exchange information, enabling them to access and use each other's functionalities.

APIs are incredibly useful because they facilitate seamless interaction between various software systems, enabling them to share data and services. They allow developers to leverage existing functionalities to create new applications or integrate different systems without having to reinvent the wheel. APIs enable quicker development, enhance interoperability, and empower the creation of innovative applications by leveraging the capabilities of other software.

There is a pre-requisite that the data ingestion is assumed to be streamlined already in the pipeline. In the following section we summarize the main features of the APIs:

A. Two Deployed Sets of APIs:

- **First API for Next day Recommendations:**

Functionality: These API offer recommendations for the upcoming day, available a day in advance, aiding proactive energy management.
Objective: To anticipate and provide suggestions for optimizing energy usage based on predictive models.
Timing: Recommendations are accessible a day prior to the specified consumption day, facilitating early access for planning.

- **Second is for Historical/Archival Recommendations:**

Functionality: Provides access to past recommendations that has been provided by the pipeline.
Objective: Serves for retrospective analysis, trend identification, and understanding energy recommendation patterns over time.

B. Priority Levels (P1 and P2):

- **P1 (Priority Level 1):** Represents critical recommendations, crucial for immediate interventions or optimizations corresponding to two lowest Integrated Scheduling Programming (ISP) scores of hours of renewable to non-renewable ratio.
- **P2 (Priority Level 2):** Offers important but less urgent recommendations compared to P1, prioritizing less critical suggestions.

C. Priority-Based Access and Usage:

- **Order of Priority:** Categorization into P1 and P2 levels provides a structured hierarchy based on urgency and importance. SMS is sent based on the priority level.
- **Usage Scenarios:** Users make decisions and allocate resources based on these priority levels, addressing critical recommendations first. The smart devices are used to validate if the user has followed up on the recommendation or not.

D. Implementation and Utilization:

- **API Access and security:** Access to the APIs, based on their requirements for energy optimization and planning is currently anonymous. However a Azure user ID is required to deploy the API from the Azure Function apps. This corresponds to a Tier 2 security layer. Since the data is anonymized, this level of security is sufficient to comply with ISO27001 protocol.

- **Interpretation of Recommendations:** Use recommendations for implementing strategies, managing consumption, and planning energy optimization activities.

E. Potential Applications:

- **Real-time Decision-Making:** Utilize P1 recommendations for immediate intervention in energy management systems.
- **Long-term Planning and Analysis:** Leverage both P1 and P2 recommendations for historical analysis, long-term planning, and identifying consistent energy consumption trends and customer behaviour.

These APIs are currently being used by Heron to test and monitor their clients electricity consumptions.

There is also a possibility to use the piplines and deploy it from packages. For that please consult the BIGG github link.

IV.2. Deployments via packages

For some use cases the pipelines can be deployed from packages. We summarize them here.

BC 1 UC 1: Energy benchmarking pipeline for buildings (link)

This use case deals with benchmarking and monitoring of energy consumption in buildings. The application can provide valuable insights for optimizing energy usage, reducing costs, and minimizing environmental impact. This contributes to more informed decision-making and improved energy efficiency in buildings.

To execute an energy benchmarking pipeline for buildings using R and MLFlow. This pipeline utilizes the BIGG Ontology harmonized data to generate results. Here's a breakdown of the steps provided in the instructions:

Software Requirements:

R: Install from the CRAN repository.

MLFlow: Install using pip (Python 3 virtual environment recommended).

BIGG AI Toolbox (biggr): Implement R package (further instructions in the link provided).

Method:

The pipeline involves a sequence of tasks including data preparation, transformation, modeling, and storage processes. This series of tasks will generate output data containing the benchmarking results for buildings.

Usage:

Input Data Requirements:

- Dataset: Example dataset (example_data.zip) aligned with the BIGG Ontology v1.0.
- Minimum Data Required:
    a. Gross floor area of the building.
    b. Hourly (or higher frequency) consumption time series.
    c. Defined energy tariffs and emissions related to the consumption time series.
    d. Hourly (or higher frequency) weather data time series.

Output Data:

Output files in Terse RDF Triple Language or in short TTL format and JSON formats containing metadata and KPI time series for buildings.

Execution Steps:

Download the Code:

Clone the repository to your local computer using Git.

```
cd <repositories_path>
git clone https://github.com/biggproject/A1-Benchmarking.git
```

Start MLFlow Server:

Start the MLFlow server from the virtual environment where it's installed.

```
cd <virtual_env_path>
```

```
source bin/activate
mlflow server --backend-store-uri sqlite:///backend.db --default-artifact-root default_artifact --host 127.0.0.1
```

Set Working Directory:

Set the working directory to the source folder of the pipeline.

```
cd <repositories_path>
cd A1-Benchmarking/source
```

Configure the config.json File:

Edit the configuration file to set the necessary parameters such as Python3 binary location, MLFlow settings, input/output data directories, etc.

Run the Pipeline:

Execute the R script Main.R.

Check Results:

Follow the logs during the execution and retrieve the results from the output folder once the process is finished.

Please note, to successfully execute this pipeline, you'll need to ensure that you have the required software installed, configure the parameters in the config.json file correctly, and provide the necessary input data in line with the BIGG Ontology v1.0. Additionally, follow the instructions provided in the linked resources for detailed installation and usage guidelines of the tools and packages mentioned.

BC 1 UC 2: Energy certification in residential buildings (link)

This pipeline is designed for Energy Efficiency Measure (EEM) assessment in buildings using Measurement & Verification techniques based on BIGG-harmonized data. Here's a breakdown of the pipeline's structure and execution steps:

Software Requirements:

R: Install from the CRAN repository.

MLFlow: Install using pip. Using a Python 3 virtual environment is recommended.

Method:

The pipeline involves a sequence of tasks including data preparation, transformation, modeling, and storage processes, leading from initial input data to the output containing assessment results.

Usage:

Input Data Requirements:

Example dataset (source/example_data.zip) includes metadata of 20 buildings that implemented EEMs and JSON files with associated time series data.

Minimum data required for each building:

Gross floor area.

Implemented EEM(s) in the building.

Hourly (or higher frequency) consumption time series.

Defined energy tariffs and emissions related to consumption time series.

Hourly (or higher frequency) weather data time series.

Output:

TTL files containing metadata for each building.

JSON files containing Key Performance Indicators (KPIs) for all buildings, stored in source/output by default.

Important Note:

Results of the assessment are provided by EEM projects, automatically grouping single EEMs that are close in time (by default, half a year). The algorithm doesn't infer energy savings produced by EEMs that are close in time.

The threshold for the auto-generation of projects is set in "EEMAssessmentConditions" fields of the source/Settings.json.

Execution Steps:

Download the Code:

Clone the repository to your local computer using Git.

```
cd <repositories_path>
git clone https://github.com/biggproject/A2-EEM-assessment.git
```

Start MLFlow Server:

Start the MLFlow server from the virtual environment where it's installed.

Set Working Directory:

Set the working directory to the source folder of the pipeline.

```
cd <repositories_path>
cd A2-EEM-assessment/source
```

Configure the config.json File: Edit the configuration file, setting parameters like Python3 binary location, MLFlow settings, input/output data directories, etc.

Run the Pipeline:

Execute the R script Main.R in the terminal.

Rscript Main.R

Check Results:

Follow the logs during the execution and retrieve the results from the output folder once the process is finished.

Ensure you've installed the necessary software, configured the parameters in the config.json file accurately, and provided the required input data aligned with the specified format and minimum requirements. Adjust configurations, inputs, and paths as needed based on your specific data and use case.

BC 4 EPC baseline (link)

This business case aims to track energy savings realized by Energy Conservation Measures (ECMs) through a before-and-after comparison of building electricity consumption. The pipeline uses machine learning to create a baseline model to predict post-retrofit consumption based on pre-retrofit data.

Input Data Requirements:

1. **Timeseries Data:**

   - Electricity consumption.

   - Outdoor temperature.

   - Global horizontal irradiation.

2. **Additional Information:**

   - Aggregation functions for time alignment.

   - Retrofit period (for model identification).

   - Building's location (for considering holidays).

Output:

- Model Identification/Training Phase:

- Model reference (e.g., MLflow URI).

- Model scores (R^2, MBE, NMBE, RMSE, CVRMSE).

- Predicted Electricity consumption timeseries.

Execution Steps:

Jupyter Notebooks:

Two Jupyter notebooks are provided to run the demo:

1. `*BC4_pipeline_training_github.ipynb*`: Trains a baseline model for electricity consumption data from harmonized data.

2. `*BC4_pipeline_prediction_github.ipynb*`: Makes predictions using the locally stored model trained in the previous notebook.

Requirements Installation:

Python and Jupyter Notebooks:

1. Ensure Python is installed on your system.

2. Install Jupyter notebooks. (https://docs.jupyter.org/en/latest/install.html).

AI Toolbox and Dependencies:

1. Install AI Toolbox:

   - As the AI Toolbox is not on PyPI, follow the installation instructions [here].

2. Install Other Dependencies:

   - Use a virtual environment if you've installed one for the AI Toolbox. Navigate to the project location and install missing dependencies using the provided `requirements.txt`.

```
pip install -r requirements.txt
```

3. Optional - Install Intel Extension for scikit-learn:

- This extension speeds up calculations for scikit-learn but requires an Intel CPU. Installation instructions are available [here](https://intel.github.io/scikit-learn-intelex/installation.html).
- To install using pip:

```
pip install scikit-learn-intelex
```

- If not installing, remove the initial lines referring to `sklearnex` in the provided notebooks.

Make sure to follow the instructions step-by-step, ensuring all dependencies are installed and the notebooks are run in the correct sequence for model training and prediction. Adjust configurations, inputs, and paths as required based on your specific use case or data availability.

BC 5 UC 11/12/13 Comfort occupants in buildings (link)

This business case and underlying use cases aim to forecast building or zone occupancy based on movement sensor data. The predicted occupancy could be utilized by various systems, such as HVAC equipment for building control or to enhance other models.

Input Data Requirements:

Timeseries Data:

Activity counter data derived from movement (IoT) sensors.

Additional Information:

Aggregation functions for time alignment.

Country information for consideration of holidays.

Output:

Model Identification/Training Phase:

Model reference (e.g., MLflow URI).

Model scores (R^2, MBE, NMBE, RMSE, CVRMSE).

Prediction Phase:

Predicted Occupancy (binary value: Occupied/Unoccupied).

Execution Steps:

Jupyter Notebooks:

Two Jupyter notebooks are provided as a demo for running this application:

BC5_pipeline_training_github.ipynb: Trains a model to learn occupancy patterns of office zones.

BC4_pipeline_prediction_github.ipynb: Makes predictions using the locally stored model trained in the previous notebook.

Requirements Installation:

Python and Jupyter Notebooks:

Ensure Python is installed on your system.

Install Jupyter notebooks. Refer to the official installation guide.

AI Toolbox and Dependencies

Install AI Toolbox:

As the AI Toolbox is not on PyPI, follow the installation instructions here.

https://github.com/biggproject/biggpy/blob/main/ai_toolbox/README.md

Install Other Dependencies:

Use a virtual environment if you've installed one for the AI Toolbox. Navigate to the project location and install missing dependencies using the provided requirements.txt.

```
pip install -r requirements.txt
```

Optional - Install Intel Extension for scikit-learn:

This extension speeds up calculations for scikit-learn but requires an Intel CPU. Installation instructions are available here.

To install using pip:

```
pip install scikit-learn-intelex
```

If not installing, remove the initial lines referring to sklearnex in the provided notebooks.

Make sure to follow the instructions step-by-step, ensuring all dependencies are installed and the notebooks are run in the correct sequence for model training and prediction. Adjust configurations, inputs, and paths as required based on your specific use case or data availability.

BC 6 UC 15 Natural Gas Demand Response (link)

This use case aims to enable a Natural Gas supplier to perform Demand Response flexibility control through the collection of consumption data from consumer buildings/apartments, by forecasting the offered flexibility of each building/home and by applying demand management algorithms based on market forecasts and energy information on top of the occupancy and home usage forecasts.

Input data requirements:

- **Timeseries Data:**

  1 month of Natural gas heating system usage data

- **Additional Information:**

  Building characteristics (house/apartment size, household size, energy class, year built, general location (climatic zone), Gas boiler Power output (kW).

Output data:

- **Action setpoints for each household**:

This includes the boiler temperature value for the corresponding house_id. The output values are sent every 1.5 minutes.

Training phase:

The model requires re-training based on new data fed to the system every 24-hours. The implementation for the demand response algorithm checks if a model from the current day is generated and available and if it is it takes it into account, otherwise the latest available model version is selected. In most implementations, the files *model.pkl* can be used.

Execution Steps:

The action coordinator code is composed in the src directory with configurations *from config.yml*. This includes a python Flask App (in *src.backend.main)* that interacts with the aggregator API, reading metrics at regular time intervals and implementing a BAU policy. To start the APP, in CMD run: python3 –m src.backend.main --host=0.0.0.0

We briefly touch on the two states the action coordinator can be in:

- Business-As-Usual (BAU): The action coordinator is in BAU when there is no DR event active. The action coordinator will send out actions to the devices every x second, where x is defined in the configuration file, please see bau_controller→control_interval_s. Now, the actions are binary, either 0 or 1, corresponding to ' off' and 'on'. When t_r is below t_r_set, the action is 1, otherwise the action is 0 for a given household.
- DR event: The action coordinator is in a DR event when there is a DR event active. The action coordinator will send out actions to the devices every x second, where x is defined in the configuration file. Please see dr_event→control_interval_s. The actions that are sent out are a direct result of the response level. A list of actions for each response level is defined at the start of each DR event. A PI-Control is used to control the response level, which would increase if the power usage were too low or decrease if the power usage is too high.

So, in both states, the action coordinator will send out actions to the devices. The action coordinator will calculate the actions based on the current state of the devices. After, it sends out the actions to the devices using the boiler aggregator API.

V. Security

V.1. Approach

The “widespread” development and usage of ICT systems are witnessing an increasing number of cyber-attacks and dynamic threats which undermines the possibilities of trusted and dependable European digital society development. Following the H2020 guidance on ethics and data protection we also respect the privacy of the personal data we process. We follow the principle of processing personal data to a strict minimum as possible. But when we do need to process it, we have secured it in the best possible way.

On the other hand, security measures cost a lot of money. To such an extent that they can undermine the financial health of a program and completely block innovative initiatives to advance our digital European community. Taking backups of data that can easily be regenerated serves no purpose, securing data that is made publicly available by the data provider is a waste of money.

Security measures are absolutely necessary, but we should only use them where they are effectively needed, and where the specific measure proves its efficiency.

For this reason, we developed a 4 step approach:

1. **Risk analysis and quantification:** For the datasets at hand across all business cases we identify the common risks at hand, affecting the data we process, building data managers and owners. We use an ISO27005[19] standardised approach and build a risk model based. The risk model identifies risks on the levels of Confidentiality, Integrity and Availability of data. Also special attention is put in the case data is personal data. The personal data impact risks identified can be used to document and demonstrate compliance with the GDPR, and (if appropriate) sectoral regulations.
2. **Define a data classification model:** Out of the input we receive from the risk analysis and data, we have grouped datasets with similar security risk profiles and create upon that a data classification model. This step allows us to generalize our approach. The data classification model is made available on the central documentation portal of the BIGG program and communicated to all stakeholders of the BIGG program In the future, a risk analysis is always done on new datasets. The risk analysis then determine which data classification level the dataset belongs to, and then immediately reuse the minimum-security measures defined for that classification level. The risk analysis is also always able to define additional security measures for specific cases.
3. **Define a minimal set of security measures:** An appropriate set of minimum-security measures was developed for each identified level of datasets in the data classification. For the determination of the security measures, we base ourselves on IS27002[20]. The measures were grouped into organizational, governance, legal and technical measures. The defined set of minimal security measures has been made available on the central documentation portal of the BIGG program and communicated to all stakeholders. It will be up to the different use case owners to implement these recommended set of security measures. Inetum has helped the use cases owners to identify and make an effort to implement these measures.
4. **Yearly monitoring and assessment of applied security measures:** Yearly monitoring the implementation of the minimal security requirements over all use cases will teach us if sufficient protection is at hand, and how feasible specific security measures are or not. This allows us to review and update the set of minimal security

[19] https://www.iso.org/standard/75281.html

[20] ISO/IEC 27002:2022(en), Information security, cybersecurity and privacy protection — Information security controls

requirements. Assessments happened across all pilots by definition of KPI's in terms of organizational, governance, legal and technical levels, which will allow pilot evaluation and replicability.

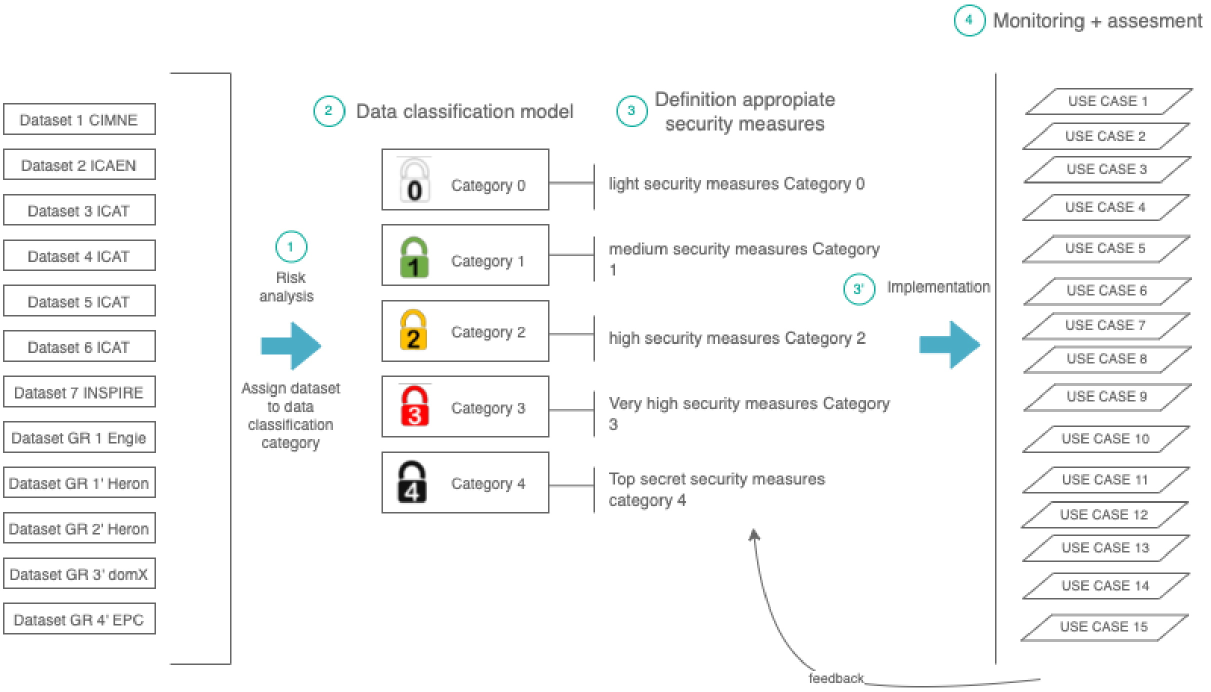
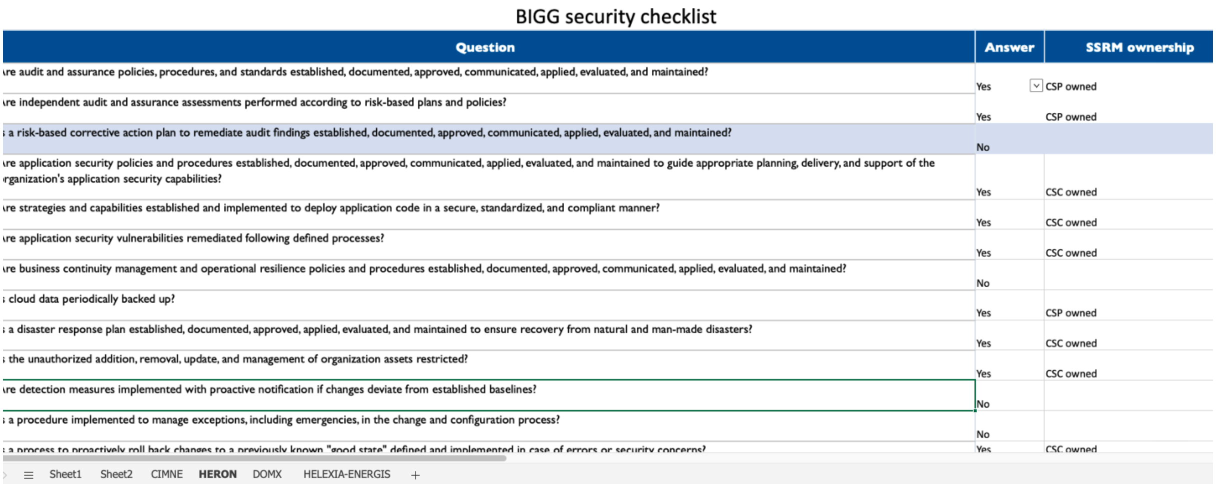


Figure 30 – Generic security approach adopted for BIGG datasets

V.2. Status

At the beginning of the project, we saw the project as a single homogenic technical solution so a questionnaire was sent to all partners to check which security measures were already in place or foreseen. This checklist was also replied by all partners. Figure 31 shows a section of the BIGG security checklist.



BIGG security checklist

Question	Answer	SSRM ownership
ıre audit and assurance policies, procedures, and standards established, documented, approved, communicated, applied, evaluated, and maintained?	Yes	CSP owned
ıre independent audit and assurance assessments performed according to risk-based plans and policies?	Yes	CSP owned
; a risk-based corrective action plan to remediate audit findings established, documented, approved, communicated, applied, evaluated, and maintained?	No	
ıre application security policies and procedures established, documented, approved, communicated, applied, evaluated, and maintained to guide appropriate planning, delivery, and support of the ırganization's application security capabilities?	Yes	CSC owned
ıre strategies and capabilities established and implemented to deploy application code in a secure, standardized, and compliant manner?	Yes	CSC owned
ıre application security vulnerabilities remediated following defined processes?	Yes	CSC owned
ıre business continuity management and operational resilience policies and procedures established, documented, approved, communicated, applied, evaluated, and maintained?	No	
; cloud data periodically backed up?	Yes	CSP owned
; a disaster response plan established, documented, approved, applied, evaluated, and maintained to ensure recovery from natural and man-made disasters?	Yes	CSC owned
; the unauthorized addition, removal, update, and management of organization assets restricted?	Yes	CSC owned
ıre detection measures implemented with proactive notification if changes deviate from established baselines?	No	
; a procedure implemented to manage exceptions, including emergencies, in the change and configuration process?	No	
; a process to proactively roll back changes to a previously known "good state" defined and implemented in case of errors or security concerns?	Yes	CSC owned

Sheet1 Sheet2 CIMNE HERON DOMX HELEXIA-ENERGIS

Figure 31 – Initial BIGG security checklist

As inputs came in and as came clear during the project that the use cases would be more heterogenous we learned that a more generic approach would be required, which could be applied in all different use cases separately easily. The ambition also grew to include a generic security framework which could be included in the BIGG framework and could be easily implemented and re-used by adopters of the BIGG framework.

Therefore, we developed and agreed with all partners the generic security approach as indicated under IV. We developed the Security management workbook and the risk register ready to identify all security related risks on the datasets across all use-cases.

Risk description						Risk calculation								
Nr	dataset ID	If	Then	Resource	Threat agent	Impact on reputatio	Legal impact	Financial impact	Personal data pr	Average impac	Highest impact	Impact index	Probability	Risk
R-001	dataset Spain 1	(integrity) Unwanted alterings could happen to the dataset during importing or during processing; so the dataset and further processing results are no longer reliable	what consequences does this have		Unwanted alterings	2- Reputation questione	1- Legal dispute	2- Max 10,000 € Acceptable cost	1 - General known ir	1,5	2	1,85	1 - occurs rarel	1,85
R-002	dataset Spain 1	(confidentiality) Due to a security breach, the dataset becomes publicly available during processing			security breach					0	0	0		0
R-003	dataset Spain 1	(availability) Suppose no backups are present. Due to a computer crash , the data in the raw data store and its modifications get lost.			Computer crash that corrupts data					0	0	0		0
R-004	dataset Spain 1	(availability) Due to certain circumstances, for example computer crashes, the data is a couple of days unavailable			Computer crash that corrupts data	1- Negligible influence (e	1- Legal dispute	1- Max 1,000 € Minimum cost	1- Max 1,000 € Minimum cost	1	1	1	2 - is rather un	2
R-005	dataset Spain 1	(Reliability) Due to security flaws in our source code, the source code of our solution can be abused for malicious goals without that we know. Malicious alterings are done to our data without us knowing								0	0	0		0
R-006	dataset Spain 2	(integrity) Unwanted alterings could happen to the dataset during importing or during processing; so the dataset and fiurther processing results are no longer reliable	what consequences does this have		Unwanted alterings					0	0	0		0
R-007	dataset Spain 2	(confidentiality) Due to a security breach, the dataset becomes publicly available during processing			security breach					0	0	0		0
R-008	dataset Spain 2	(availability) Suppose no backups are present. Due to a computer crash , the data in the raw data store and its modifications get lost.			Computer crash that corrupts data					0	0	0		0
R-009	dataset Spain 2	(availability) Due to certain circumstances, for example computer crashes, the data is a couple of days unavailable								0	0	0		0
R-010	dataset Spain 2	(Reliability) Due to security flaws in our source code, the source code of our solution can be abused for malicious goals without that we know. Malicious alterings are done to our data without us knowing								0	0	0		0
R-011	dataset Spain 3	(integrity) Unwanted alterings could happen to the dataset during importing or during processing; so the dataset and fiurther processing results are no longer reliable			Unwanted alterings					0	0	0		0
R-012	dataset Spain 3	(confidentiality) Due to a security breach, the dataset becomes publicly available during processing			security breach					0	0	0		0
R-013	dataset Spain 3	(availability) Suppose no backups are present. Due to a computer crash , the data in the raw data store and its modifications get lost.			Computer crash that corrupts data					0	0	0		0
R-014	dataset Spain 3	(availability) Due to certain circumstances, for example computer crashes, the data is a couple of days unavailable								0	0	0		0
R-015	dataset Spain 3	(Reliability) Due to security flaws in our source code, the source code of our solution can be abused for malicious goals without that we know. Malicious alterings are done to our data without us knowing								0	0	0		0
R-016	dataset Spain 4	(integrity) Unwanted alterings could happen to the dataset during importing or during processing; so the dataset and fiurther processing results are no longer reliable			Unwanted alterings					0	0	0		0
R-017	dataset Spain 4	(confidentiality) Due to a security breach, the dataset becomes publicly available during processing			security breach					0	0	0		0
R-018	dataset Spain 4	(availability) Suppose no backups are present. Due to a computer crash , the data in the raw data store and its modifications get lost.			Computer crash that corrupts data					0	0	0		0
R-019	dataset Spain 4	(availability) Due to certain circumstances, for example computer crashes, the data is a couple of days unavailable								0	0	0		0
R-020	dataset Spain 4	(Reliability) Due to security flaws in our source code, the source code of our solution can be abused for malicious goals without that we know. Malicious alterings are done to our data without us knowing								0	0	0		0
R-021	dataset Spain 5	(integrity) Unwanted alterings could happen to the dataset during importing or during processing; so the dataset and fiurther processing results are no longer reliable			Unwanted alterings					0	0	0		0
R-022	dataset Spain 5	(confidentiality) Due to a security breach, the dataset becomes publicly available during processing			security breach					0	0	0		0
R-023	dataset Spain 5	(availability) Suppose no backups are present. Due to a computer crash , the data in the raw data store and its modifications get lost.			Computer crash that corrupts data					0	0	0		0
R-024	dataset Spain 5	(availability) Due to certain circumstances, for example computer crashes, the data is a couple of days unavailable								0	0	0		0
R-025	dataset Spain 5	(Reliability) Due to security flaws in our source code, the source code of our solution can be abused for malicious goals without that we know. Malicious alterings are done to our data without us knowing								0	0	0		0
R-026	dataset Spain 6	(integrity) Unwanted alterings could happen to the dataset during importing or during processing; so the dataset and fiurther processing results are no longer reliable			Unwanted alterings					0	0	0		0
R-027	dataset Spain 6	(confidentiality) Due to a security breach, the dataset becomes publicly available during processing			security breach					0	0	0		0
R-028	dataset Spain 6	(availability) Suppose no backups are present. Due to a computer crash , the data in the raw data store and its modifications get lost.			Computer crash that corrupts data					0	0	0		0
R-029	dataset Spain 6	(availability) Due to certain circumstances, for example computer crashes, the data is a couple of days unavailable								0	0	0		0
R-030	dataset Spain 6	(Reliability) Due to security flaws in our source code, the source code of our solution can be abused for malicious goals without that we know. Malicious alterings are done to our data without us knowing								0	0	0		0
R-031	dataset Spain 7	(integrity) Unwanted alterings could happen to the dataset during importing or during processing; so the dataset and fiurther processing results are no longer reliable			Unwanted alterings					0	0	0		0
R-032	dataset Spain 7	(confidentiality) Due to a security breach, the dataset becomes publicly available during processing			security breach					0	0	0		0
R-033	dataset Spain 7	(availability) Suppose no backups are present. Due to a computer crash , the data in the raw data store and its modifications get lost.			Computer crash that corrupts data					0	0	0		0
R-034	dataset Spain 7	(availability) Due to certain circumstances, for example computer crashes, the data is a couple of days unavailable								0	0	0		0
R-035	dataset Spain 7	(Reliability) Due to security flaws in our source code, the source code of our solution can be abused for malicious goals without that we know. Malicious alterings are done to our data without us knowing								0	0	0		0
R-036	dataset Greece 1	(integrity) Unwanted alterings could happen to the dataset during importing or during processing; so the dataset and fiurther processing results are no longer reliable			Unwanted alterings					0	0	0		0
R-037	dataset Greece 1	(confidentiality) Due to a security breach, the dataset becomes publicly available during processing			security breach					0	0	0		0

Datasets | RiskRegister | TechnicalSolutionUseCases | +

Figure 32 – Risk register

The comprehensive risk register is followed by the translation of identified risks into specific security levels for our dataset using the generic security approach shown in Figure 32. To operationalize this strategy, we have developed a detailed security control checklist featuring controls categorized across various domains such as access control, operations security, and more, each aligned with security levels ranging from 0 to 2 as defined in our overarching security approach. The responsibility for implementing these controls rests with our partners, who are entrusted with completing the checklist and providing thorough explanations for any aspects not implemented. This structured approach not only fosters accountability but also ensures a systematic and collaborative effort to mitigate risks associated with our dataset. We anticipate that this process will not only enhance the overall security but also facilitate a continuous improvement cycle through regular reviews, feedback loops, and a commitment to adaptability in response to evolving security landscapes. Table 1 gives the domain and category wise split between control measures.

Table 1 – Summary count of the control measures based on domain and category level.

Domain	Category 0	Category 1	Category 2
Access control	7	6	5
Asset management	1	1	4
Business continuity management	0	3	0
Communications security	3	0	0
Compliance	2	0	1
Cryptography	3	0	8
Incident management	4	0	0
Information security incident management	1	3	0
Information security policy	1	0	3
Operations security	15	7	15
Organization of information security	2	2	1
Physical and environmental security	4	1	1
Safe staff	0	0	3
Supplier relationships	0	1	2
System acquisition, development and maintenance	0	2	1
Total control measures	**43**	**26**	**44**

V.2.1. Key items from partners on the security controls

The Security control measure checklist distribution process was successfully executed, with the responsible partners having received and acknowledged the checklist. During this phase, clear communication was established to ensure partners comprehended the significance of each security control measure and its corresponding security level, along with explicit expectations and deadlines for completion. Subsequently, partners were diligently requested to furnish comprehensive document for each implemented security control measure, accompanied by explanations for any instances where controls were not put in place. The document emphasized the risk mitigation strategies employed and highlighted how the implemented security control measures elevate the overall security level of the dataset. Following the documentation phase, a robust review and validation process were implemented to assess the completeness and efficacy of the security controls. Internal security experts were engaged in scrutinizing the controls to ascertain their appropriateness in addressing identified risks. Feedback based on the review results has been provided to the partners, and collaborative efforts have been established to address any identified gaps or areas for improvement. These iterative feedback loops are integral to refining and enhancing the security measures over time, ensuring an adaptive and resilient security framework. Below we summarize the main feedbacks that was recieved fromparteners on implementations of the security measures control per domain.

Table 2 – Main highlights of domain wise implementation performed by respective partners

Domain	CIMNE	Energis	InetumRD (IRD)	DomeX
Use case	**UC 1,2,3,4,5,6,7**	**UC 8,9,10,11,12,13**	**UC 14**	**UC 15**
Access control	The user set his/her password by using email verification, CIMNE don't communicate passwords	-- Using email address and password. -- Passwords should have a minimum length of 12 characters and contain at least 1 special character and at least 1 number.	-- Password strength: length: 12 characters, requirement to change password on first login, comprising lower and upper case letters, numerals and special characters. -- Personal accounts that have not been used for more than 6 months are to be disabled automatically	-- Users have access only to the resources that have been assigned by administrators. -- Password requires length of minimum of 8 characters, and 1 Number 1 Capital letter and change every 1 year.
Asset management	CIMNE provides an extraction of all data and will remove all data when necessary.	Energis centralizes all the data needed within the framework of its assignment in a secure data center in accordance with applicable industry standards. Physical access to the data center is monitored and managed.	All digital data that is processed by Inetum-Realdolmen within the framework of its assignment are stored centrally in the data center in an encrypted fashion on the hard drives.	Handled by the cloud provider.
Business continuity management	The backups are stored in back up cloud applications and are not handled. CIMNE can recover by using the daily backups.	Daily backups of the filesystems is implemented which can be redeployed in case of disaster. A full backup of the Energis.Cloud environments is performed every night. An alarm is generated in case a backup could not be made.	Inetum-Realdolmen has an emergency recovery plan in case of disasters relating to its servers in the data center on which personal data is stored.	It is implemented as PoC.
Communications security	CIMNE protects administration flows by using HTTPS protocols	HTTPS and SFTP are used as much as possible. Certificates are purchased via Combell and Let's encrypt.	All data sent over public networks is encrypted.	-- Oauth + HTTPS -- Application is developed in a VPC.
Compliance	Security compliance implemented.	Security vulnerability tests are made every month or with major Energis Cloud releases by Wazuh IDS open-source tool.	Baseline corporate tools (see document "minimal security norm for suppliers").	Security compliance implemented.
Cryptography	CIMNE use *letsencrypt* application to generate all certificates.	The requirements for the protection of internal data and systems that Energis uses to provide its services are analyzed and identified in collaboration with our IT supplier(s).	The requirements for the protection of internal data and systems that Inetum-Realdolmen uses to provide its services are analyzed and identified in collaboration with our IT supplier(s).	All data are encrypted.
Incident management	CIMNE maintains a register of security breaches with a description of the breach, the time, the consequences of the breach, the name of the reporting person and the person receiving the report.	Security events are considered as critical incidents and treated accordingly based on our SLA.	Security updates are responded to and installed in accordance with its patch management process.	All of incidents is handled by internal team.

Information security incident management	CIMNE has internal procedures.	All incoming data are anonymized by the provider. So GDPR compliant.	All incoming data are anonymized by the provider. So GDPR compliant.	From development to implementation is performed.
Information security policy	CIMNE in case of disaster has an emergency recovery plan to recover a previous state.	All incoming data are anonymized by the provider. So GDPR compliant.	All incoming data are anonymized by the provider. So GDPR compliant.	DomeX has a data processor agreement with the processor.
Operations security	All new versions must pass a list of tests including security tests.	Thanks to the SaaS architecture, there is one single version of Energis cloud infrastructure.	IRD carries out anti-malware checks on all its internal systems to prevent malicious software from obtaining unauthorized access to customer data.	All aspects implemented (except yearly restoring test).
Organization of information security	CIMNE IT experts control and form the new employees with privacy and security policy rules and procedures. All security controls are checked and reported monthly.	This is part of Energis security policies.	IRD has appointed a Data Protection Officer/Privacy Officer to supervise compliance with the Belgian Privacy Act and the EU General Data Protection Regulation (GDPR) within Inetum-Realdolmen.	Except monthly reporting, all aspects like review etc. are implemented.
Physical and environmental security	Third-party subcontracted. The application is hosted in OVH infrastructure.	This is done by Energis' data center provider.	IRD centralizes all the data needed within the framework of its assignment in a secure data centre.	Not performed.
Safe staff	CIMNE has assigned a contact and information security coordinator.	This is part of Energis' security policies.	Employees are informed of any significant changes in this regard by email.	Not performed
Supplier relationships	Not required as currently this is PoC.	Not required as currently this is PoC.	Not required as currently this is PoC.	Not required as currently this is PoC.
System acquisition, development and maintenance	CIMNE has introduced a segregation system to ensure that persons do not obtain access to data they do not need for carrying out their job responsibilities.	This is part of the Security Development Lifecycle (SDL).	It is checked (for the role) when new users are added with certain IAM role.	Not required as currently this is PoC.

V.2.2. Why Category 3 or 4 level measure was not needed in BIGG?

The category 3 or 4 level security was not necessary for BIGG because of the following two reasons:

- **Nature of Data:** Our use case primarily involves IoT data from sensors in building, meter or appliances and weather data. Unlike Category 3 and 4 security levels, which typically deal with highly sensitive and top-security-level datasets, our data falls into a less sensitive category.
- **Anonymization**: Category 3 and 4 security measures are designed for situations where the utmost precautions are necessary due to the sensitivity of the data involved. However, in our context, the nature of the data being either anonymized or publicly accessible. For data anonymization, the responsibility lies primarily with the data provider. When possessing

customer data to make it available to other partners and associated details, they undertake the crucial task of anonymizing the information by assigning pseudo numbers, texts, or a combination thereof to any elements containing identifiable personal data. This meticulous process ensures that individual identities remain obscured, safeguarding privacy and aligning with data protection regulations. Once anonymized, the data is made available to the relevant entities. In cases where the data remains within the confines of the provider's own tenant, internal processing occurs with strict adherence to robust security measures and compliance with GDPR guidelines—parameters that customers have consented to, thereby reinforcing a commitment to data security and privacy. Thus, it makes the implementation of Category 3 and 4 top-level security measures unnecessary.

V.2.3. Conclusion about security aspects and its future

We project our future development of European digital society in a strong and robust way, but we choose to implement effective and efficient measures against the specific risks which are there. We make the balance between effective protection but also allowing space for innovation and further develop our European region in a sustainable way.

We developed a generic security approach, data classification model, as well as the recommended minimal set of security measures for each classification level. The security questionnaire is available in the BIGG total solution in GitHub (link)[21] to that future adopters of the BIGG framework so that anyone can easily re-use this security approach.

However, looking ahead, future considerations must be made for data security to stay abreast of evolving technologies and advancing data handling practices. As we navigate an ever-shifting digital landscape, these three key points can serve as guiding principles to ensure that security measures remain not only robust today but also agile and responsive to the challenges that lie ahead. These are:

1. **Adaptive Anonymization Techniques:** As technology continues to evolve, data providers should explore and adopt adaptive anonymization techniques that can effectively counter emerging threats and vulnerabilities. This may involve leveraging artificial intelligence and machine learning algorithms to dynamically adjust anonymization methods based on evolving patterns of data breaches.
2. **Integration of AI and Machine Learning security layer:** Embrace artificial intelligence (AI) and machine learning (ML) technologies in security measures. These technologies can play a crucial role in identifying patterns, anomalies, and potential security threats. Implement self-learning algorithms that can adapt to new attack vectors and evolving cyber threats over time.
3. **Continuous Monitoring and Updating**: Establish a dynamic security framework that involves continuous monitoring and regular updates. This includes staying informed about emerging threats, vulnerabilities, and technological advancements. Implementing a system for real-time threat intelligence feeds and automatic updates helps in proactively addressing potential risks.

To establish a future-proof security framework, organizations must prioritize adaptability and scalability, ensuring that their defenses can seamlessly evolve alongside emerging challenges.

[21] https://github.com/biggproject/Demo-security-control-measures/tree/main

VI. CONCLUSION AND FUTURE PROSPECTS

In the BIGG project we have managed to create a solid, dynamic solution architecture. This architecture has been based on discussions between the different teams and work packages. The described use-cases have serve as a validation for the architecture. Adjustments can be made when necessary if the use cases are to be deployed in a different business context, thus the architecture have demonstrated its flexibility.

We have spent time aligning all the work packages and partners. This alignment includes boundaries between the work packages, the architecture, supported data formats and the responsibilities for each component and the security standard to adhere.

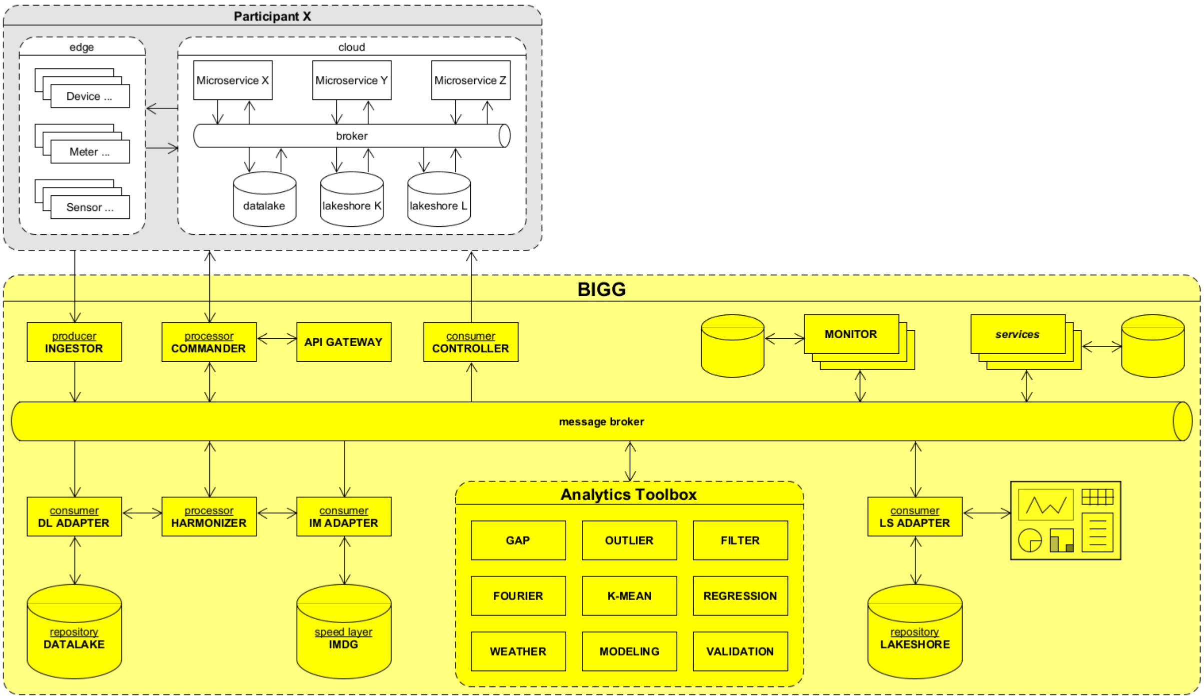


Figure 33 – BIGG Reference Architecture

Figure 33 gives the reference architecture that is common to all the use cases. Yellow blocks in Figure 33 are the components that form the BIGG Reference Architecture Framework (RAF) and work Package 3 has developed and securely deployed several of these blocks.

To summarize and conclude the deliverable of WP3, the Communication Layer from Task 3.1 handles ingesting, and processing data. We have identified that the communication layer plays a crucial role in creating the components needed to handle different types of building data. Following a state-of-the-art RAF designed in WP2, we have created special components (Ingestor) to support various protocols like HTTP and MQTT for raw data ingestion. In Work Package 3 the different business cases have derived their own custom ingestor components, based on their needs of processing the data. Processing can mean storing in a BIGG data store, handing it over to the AI Toolbox for analysis or pushing it onto the message broker. The raw data is then pushed onto the BIGG streaming bus system, intercepted, and stored in dedicated raw-data stores. It triggers harmonization components to transform the raw data into a standardized format following WP4 specifications. The harmonized data is published on the messaging system, stored in dedicated stores, and can be intercepted by AI toolbox components. These AI tools process the data to activate specific pipelines, creating knowledge from it. The resulting insight data is broadcasted within the BIGG system and can be intercepted by dashboard controllers, formatting it for custom dashboards, or external system controllers, pushing it back to specific external systems. The modularity of BIGG ensures high reusability, adapting communication components to meet specific customer requirements. We have created graphical user interfaces in the form of dashboards to monitor pilot sites and user applications per use case. There are regular interactions with the customers on their comfort and ease of usage. The current interfaces have thus evolved to its present form after several iterations keeping in mind the ease of usage of the customers.

In addition to that end-user applications, both graphical and non-graphical platforms, were developed to provide solutions for various data-based scenarios, integrating innovative services into existing systems or creating new ones. We also provide extensive guidelines for replicating use cases through API or manual deployment, along with specifications for necessary setups to give full overview of the solution. The final task in this use case was to provide guidelines on how to securely integrate and deploy the BIGG components in clients infrastructures. The Security measures and risks were assessed, and partners committed to adhering to ISO27001 standards by filling a comprehensive security measures template. Adhering to ISO27001 security standards is a resolute commitment that significantly enhances the overall reliability and trustworthiness of the BIGG. Through comprehensive risk management, data confidentiality measures, robust governance practices, resilient infrastructure security, and a culture of continuous improvement, BIGG partners demonstrate their dedication to safeguarding critical assets and information. It also ensures that the BIGG components operate with transparency and commitment to adhere to the legal requirements.

The future of an open-source project like BIGG dedicated to creating an ontology for interoperability and piloting applications in the energy sector appears promising. As the project gains traction and proves its efficacy, BIGG partners look forward for a widespread adoption within the energy industry, setting the stage for potential applications in related sectors. The project's open-source nature fosters a collaborative community, facilitating ongoing enhancements and adaptability to emerging technological trends. Success in the energy sector could position the project as a key influencer in shaping interoperability standards, potentially leading to broader industry-wide adoption among all other EU projects that have shared goals. The project's impact could extend beyond its initial scope, finding applications in diverse domains facing similar interoperability challenges. Additionally, the project's success may attract partnerships, increased funding opportunities, and recognition from policy and regulatory bodies, further ensuring its sustained relevance and impact.

BIGG