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 preliminary end-user communication and security layers

Deliverable 3.1

Responsible:	Nico Vermeir (Inetum-Realdolmen)
Document Reference:	D3.1
Dissemination Level:	Public
Version:	1.0
Date:	22nd March 2022

Executive Summary

BIGG, or Building Information aGGregation, is an European Commission H2020 project (https://www.bigg-project.eu/) 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 is the first deliverable of Work Package 3. As a reader of this document you will read about the purpose of Work Package 3, its current status and the scheduled work up until the next deliverable.

In this first phase of the project we have aligned our architecture through the 6 business cases. The business cases have started developing ingestion components that will be used as a base line for generic components in the next phase of the project.

Contributors Table

<table>
<tr><th>Document Section</th><th>Author(s)</th><th>Other Contributors to results</th><th>Reviewer(s)</th></tr>
<tr><td>Introduction</td><td>Nico Vermeir (Inetum BE)</td><td>-</td><td rowspan="5">Michel Hours (Inetum FR)
Stoyan Danov (CIMNE)</td></tr>
<tr><td>Communication Layer</td><td>Jarne Kerkaert, Nico Vermeir (Inetum BE)
Nicolas Pastorelly (CSTB)
Jordi Carbonell (CIMNE)
Riccardo Devivo (Helexia - Energis)</td><td rowspan="3">Michel Hours (Inetum FR)
Jordi Cipriano, Daniel Pérez, Xavier Durán, Carolina Sareyan (CIMNE)
Eric Pascual (CSTB)
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)</td></tr>
<tr><td>Graphical applications</td><td>Jarne Kerkaert, Nico Vermeir (Inetum BE)
Jordi Carbonell (CIMNE)
Pierre Lehanneur (HELEXIA)
Polychronis Symeonidis (DomX)</td></tr>
<tr><td>Security</td><td>Veerle Van Keyenbergh, William Debruyn (Inetum BE)</td></tr>
<tr><td>Conclusion</td><td>Nico Vermeir (Inetum BE)</td><td>-</td></tr>
</table>

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
HVAC	Heating, ventilation, and air conditioning
AITB	BIGG AI Toolbox
RAF	Reference Architecture Framework

I. INTRODUCTION

I.1.1. Purpose and structure of the document

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
- Task 3.2: Graphical User Interfaces Development
- Task 3.3: Security Layer Development

This document corresponds to the first deliverable 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, current status and next steps.

The document is structured in the same order as the task definitions. After this introduction chapter the document describes the Communication Layer (section II), the Graphical Interfaces (section III) and the Security layer (section IV).

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. 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. 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 will ingest and 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 D2.2), the communication layer components are laying in the northbound and in a southbound of the architecture as presented in the following simplified schema:

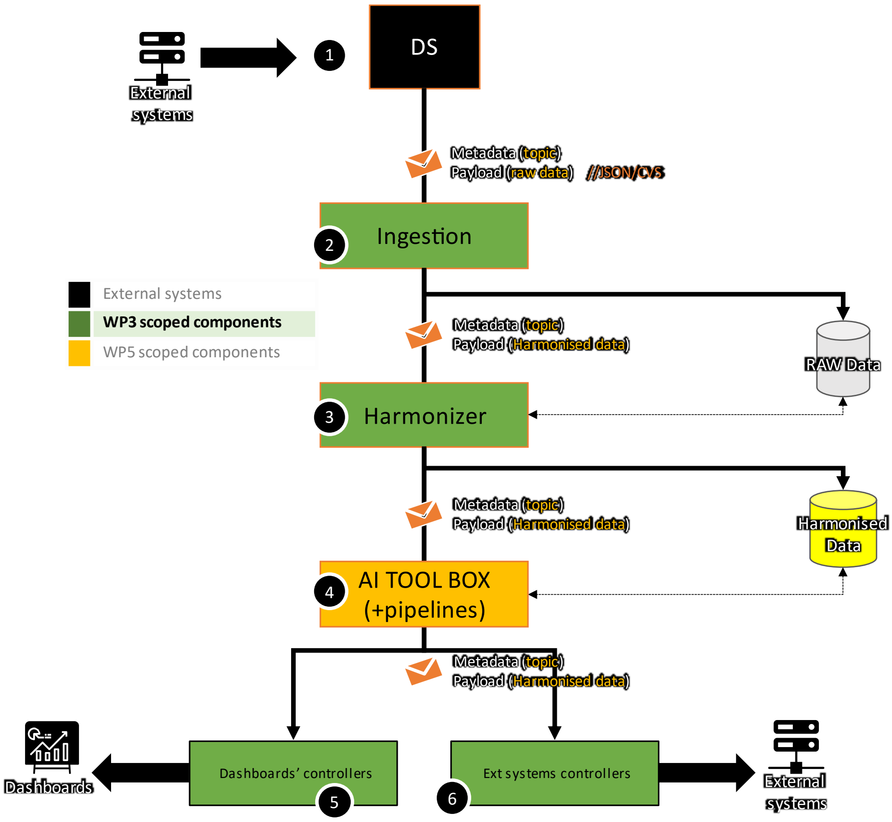


Figure 1 - Simplified communication layer architecture

Next to the Northbound of the BIGG architecture, technical services of external systems are responsible of exposing endpoints to inject datasets (see 1 in Figure 1) 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 Figure 1) have been designed to support various protocols (HTTP, MQTT etc.). 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 Figure 1). 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 will demonstrate the potential of BIGG for the collection, harmonisation and analytical processing of a large number of datasets belonging to different departments of the Catalan administration.

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[1]

II.2.1.a. Status

Sources import process

In this phase of the project 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[2] Database with all Spain electricity consumptions.

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

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

- Weather data for all Europe stations. Provider: DARKSKY[3].
- 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[4].
- Billing information systems. Monthly and hourly gas and electricity consumption data. Provider: GENWEB[5].

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)
- Energy certification data for public and private sector buildings. Provider GENCAT OpenData system. Information about 1.000.000 BPCertifiates of private and public buildings

<u>Ingestion process</u>

API clients and file parsers are implemented via scripts, created ad-hoc for each resource. The ingestion process oversees collecting the data using the provider's instructions and send it to a message broker's topic (Kafka[6]).

<u>Link data and harmonization process</u>

The harmonization is performed by reading the data from the message broker's topic, identifying which is the source that sent it, and applying a custom script to transform the data to triples in RDF[7], following the BIGG data model developed in WP4 (see deliverable D4.1).

This process has to be done ad-hoc for each source.

<u>Data storage</u>

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, previous to harmonization. This is done to keep a copy of the gathered data in case we need to process it again. These BCs use HBASE[8] 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[9] 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.

[3] https://darksky.net/

[4] https://www.ixon.cloud/

[5] http://www.gemweb.es/

[6] https://kafka.apache.org/

[7] https://www.w3.org/RDF/

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

[9] https://neo4j.com/

Presentation

The presentation of the data in the current version is done through a User Interface developed ad-oc to cover the BC 1, 2 and 3 user's needs. The current 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). See section III.1.2.a. for more information.

The presentation of the results of the AI toolbox, still under development, is not implemented in this version of the User interfaces. In this phase, the interaction with the AI toolbox is done through a Jupiter notebook that allows to test the first results of the analysis.

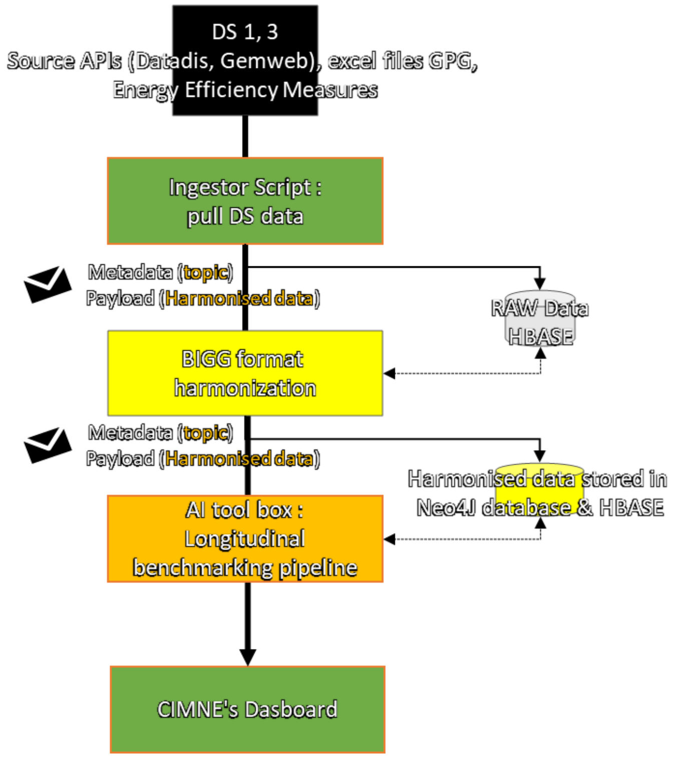


Figure 2 – Communication Layer Architecture

II.2.1.b. Next step

The next steps in the implementation of business cases 1, 2 and 3 will be:

Source import process

- Improve current resources connected
  - In this phase, the quality of the input data will be improved, as the end-users already have access to it. This will improve the linking of data between them, in cases where this could not be done automatically, and users will also be able to provide new data and verify the data they have available.
- Connect new data resources such as:
  - CMMS systems (mantest, waiting for provider to enable API connection).
  - Cadastral information (API implemented in test mode).
  - BIM models (developing the BIM file parser).
  - Energy monitoring systems (DeXMA, Api client developed, waiting for credentials).

Ingestion process

- Improvement of the current ingestion process
  - Scale server nodes with data volume needs

- Set the ingestion for the new resources connected

Link data and harmonization process

- Use standard tools to harmonize and transform the data to the BIGG Ontology such as RML[10], sparql-generator[11] or Yarn[12].
- Creation of a user-friendly tool to generate the mappings without the need to understand and know the ontology developed in WP4.

Presentation

- Implementation in the final User's Interface the results of the AI toolbox for the BC 1, 2 and 3 pipe lines developed in WP5

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 Engie company. 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 modules to be implemented must offer:

- Inputs
  - Data extracted from BMS, IoT sensors (consumptions, temperatures, occupancy, etc.), utility invoices, status and forecast from weather services
- Computations
  - Relevant KPIs which can be viewed in dashboards and reports
  - Identification of baseline models with dependencies from influencing factors such as weather and occupancy (IPMVP[13])
  - Optimisation service module addressing the different goals related to energy efficiency, renewable energy usage, occupancy, comfort and cost of the building
- Outputs

[10] https://rml.io/specs/rml/

[11] https://www.w3.org/TR/rdf-sparql-query/

[12] https://yarnpkg.com/package/json-transforms

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

- Predictive consumptions model to compare actual with forecasted data
- Optimization parameters for the rule engine of the Controller module to pilot the building appliances (HVAC)
- Dashboards and reports

II.2.2.a. Status

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

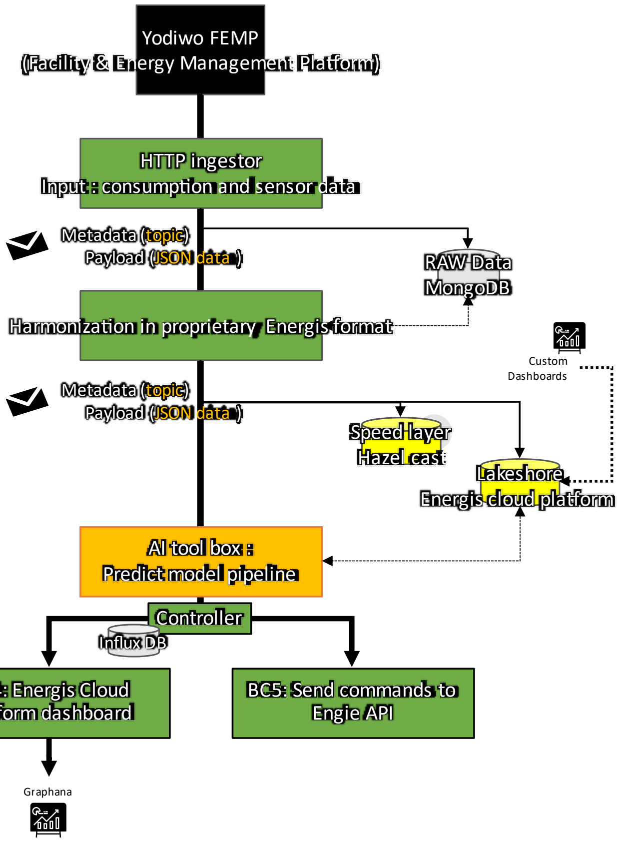


Figure 3 – Ingestor Pipeline

- 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 III.1.2.b. - HELEXIA User Interfaces 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 forecasting model producing a CSV manually uploaded to Energis.cloud. This output is used for comparison purpose between the AITB model and the model produced by the Energis.cloud internal IPMVP module.
- 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 Engie Connect service to tune its internal rule engine and, in turn through Engie 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 in section III.1.2.b. - HELEXIA User Interfaces below.

II.2.2.b. Next step

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

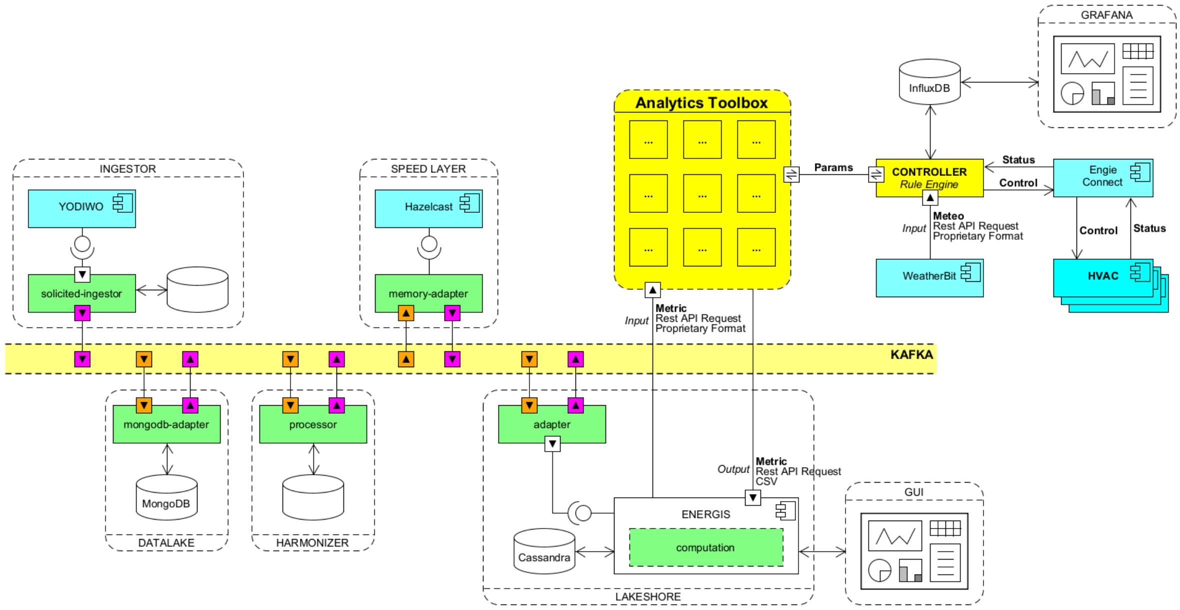


Figure 4 – BIGG Architecture Diagram

This is the list of activities to enact to completely fulfil the requirements exposed in the beginning part of this section:

- Harmonization
  All the different components must use the BIGG RDF ontology format instead of proprietary JSON format
- Integration
  The Analytics Toolbox must be integrated with the platform architecture in terms of communication (e.g., using Kafka topics instead of direct REST API calls) and must be wrapped by a microservice exposing its main functionalities.
  The "memory-adapter" microservice must be more and more used by the other modules requiring fresh data in near real-time, moreover it could be used as a common cache to store relevant entities and metrics shared between all the platform components.
- Automation
  Without the user intervention (launching script, uploading CSV results, etc.) the system must work independently and autonomously. The Analytics Toolbox must be

continuously collecting data to enrich and tune its predictive model to produce more and more accurate forecasts. On the other hand, Energis.cloud platform must be instantly populated with results coming from the AITB to show dashboard and compile reports for monitoring purposes.

- 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 could be persisted into MongoDB as a document to be in a second moment retrieved to speed up AI and ML computation procedures. Furthermore, collecting models in different seasons could be useful to develop a genetic selection algorithm to discover the most accurate model for the period in the year.
- Implementation
  The WeatherBit service is accessed by Energis.cloud too, so it would be 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 could 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 must be 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 could be summarized in the following to-be architectural diagram:

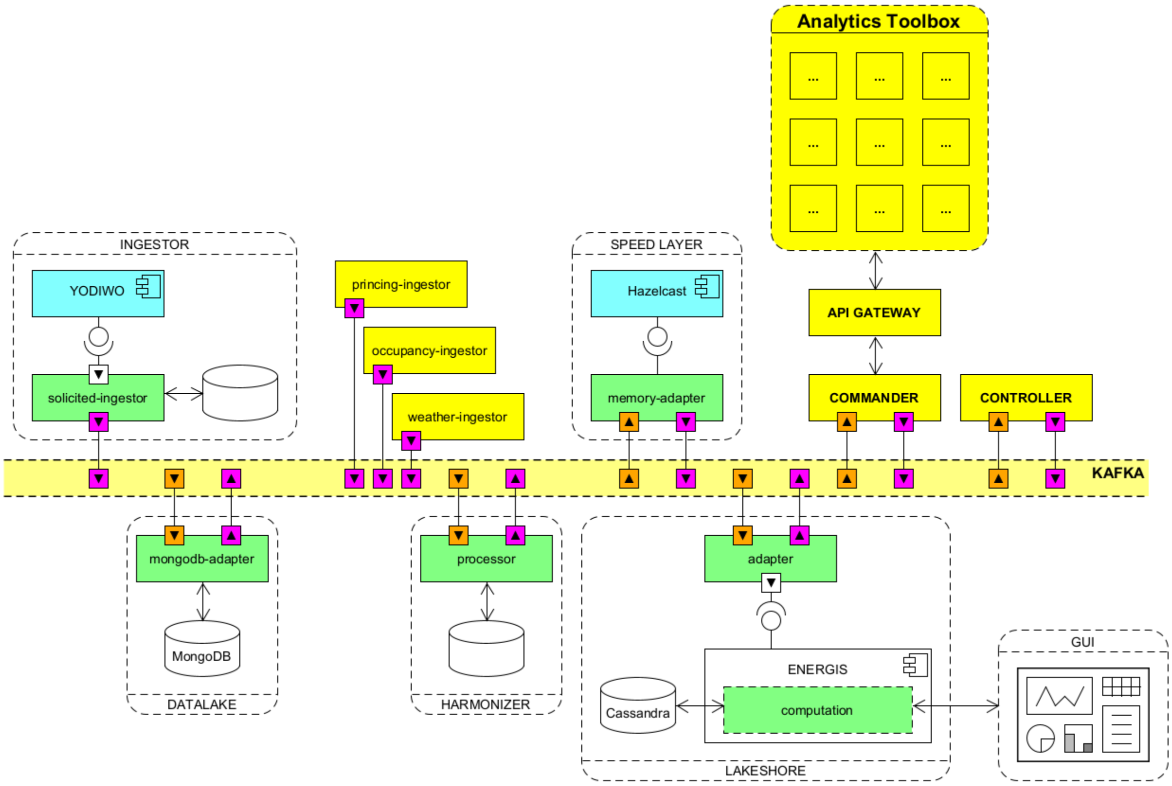


Figure 5 – BIGG Architecture Diagram

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
- Datase 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.) will be made available for offline access, to assist the building performance analysis.

The aforementioned 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.

II.2.3.a. Status

UC#14 and UC#15

The following flow diagram describes the initial 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.

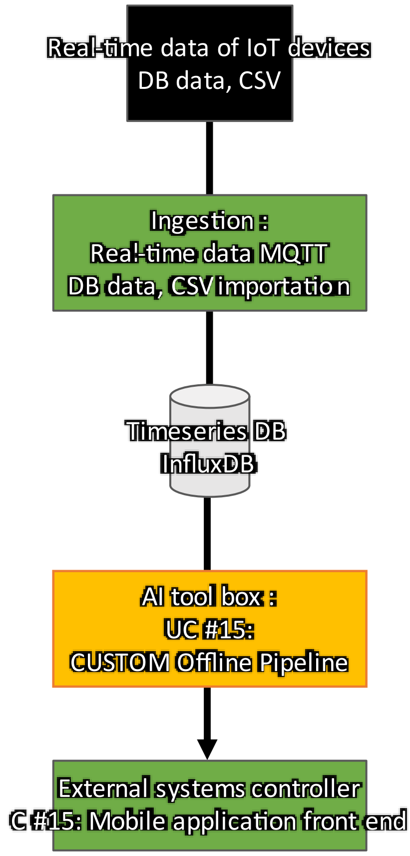


Figure 6 – Components used in UC15

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 III.1.2.d. - DomX User Interfaces.

II.2.3.b. Next step

The next steps of the 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.

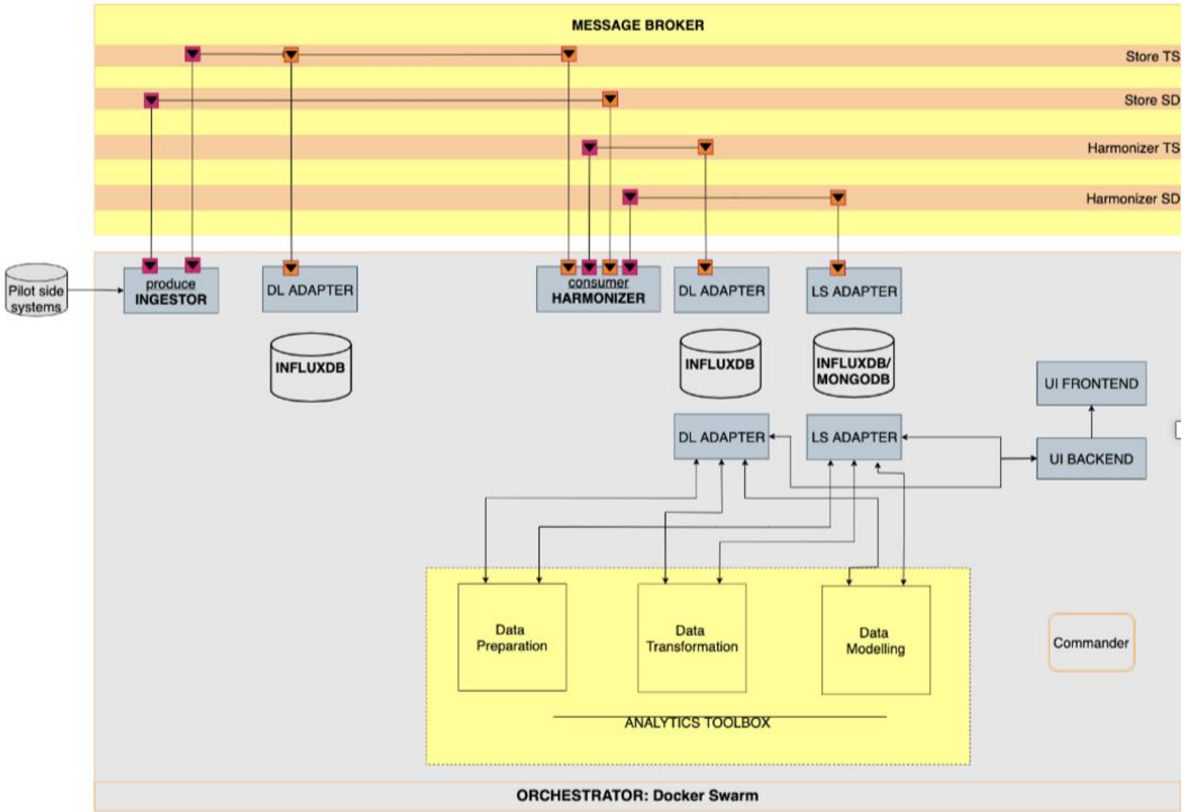


Figure 7 – BIGG Architecture

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

Apache Kafka:

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

Ingestion:

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

Harmonization:

The Data Harmonization is a big part of the BIGG architecture and we plan to use the BIGG Harmonizer component implemented which will transform raw data collected from the datasources involved in BC#6 and output data in the BIGG Format ready to be used by other components of the pipeline.

AI Pipeline:

The current AI Pipeline will be changed to include 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. More specifically,

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 is that the BIGG AI Pipeline will receive harmonized data through KAFKA topics and will perform analytics and produce meaningful insights in an automated way.

External System Controller:

Improvements will be made to the External System Controllers. DomX's Mobile App will be 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 will be 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.

III. GRAPHICAL INTERFACES

III.1. Status

A key characteristic of the BIGG solution is its modularity. It guarantees a high rate of general reusability. In order to ensure the fulfilment of the business case-oriented approach of BIGG, end-user applications (both graphical and non-graphical) will be 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 will 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 new ones (e.g. Mobile apps, monitoring tools). To this end for Task3.2 it was decided to develop:

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

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 will visualize 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.

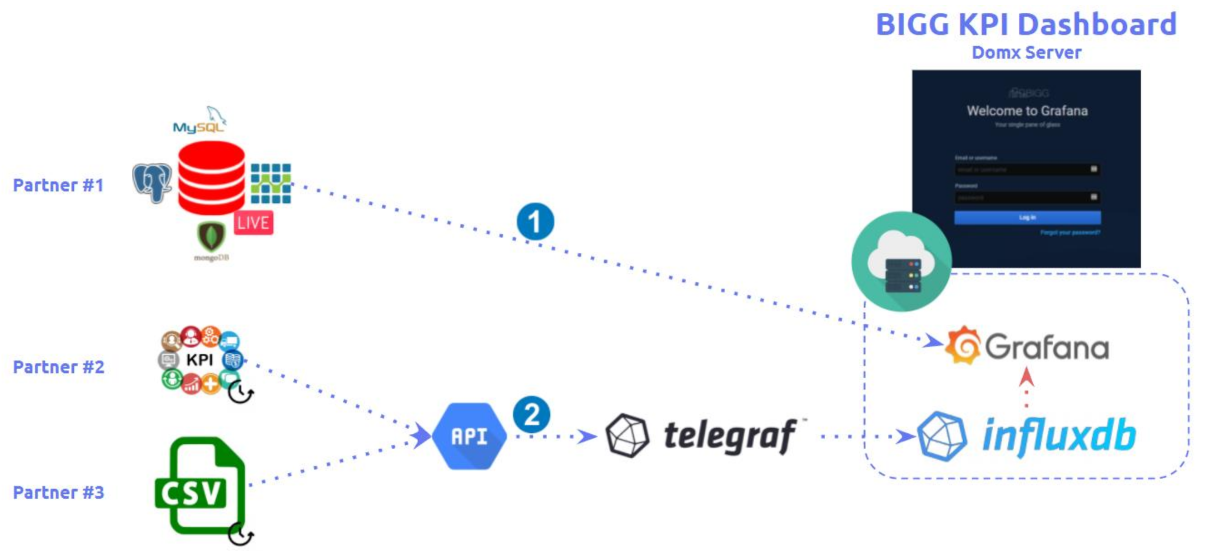


Figure 8 - BIGG Pilot Monitoring Dashboard Architecture

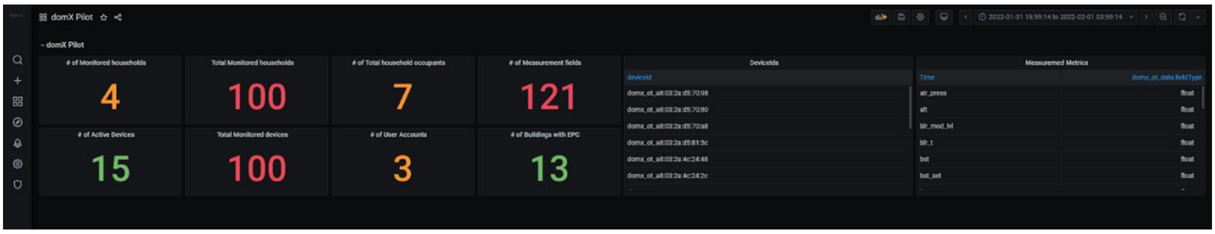


Figure 9 - Proof-of-Concept Implementation for the BIGG Pilot Monitoring Dashboard (Grafana)

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

III.1.2. End-user applications & dashboards

III.1.2.a. CIMNE User Interfaces

A new web application is being developed for the users of business cases 1, 2 and 3. The development of the user interface has been divided into two phases aligned with the project. The specification and initial functional design have been prepared and is being implemented.

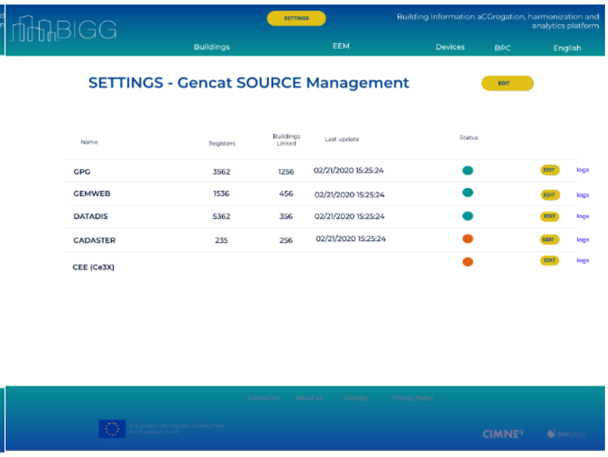
The first phase of the user application, which includes the management and presentation of the input data, is currently in progress. The current development focuses particularly on linking the various data resources to each other; facilitating the creation, modification and correction of input data; and user management and building organization.

This first phase is intended on the one hand to start involving the end users to validate the BIGG developments and on the other hand to ensure and verify the consistency of the input data so that they are available for the analytical developments of the project.

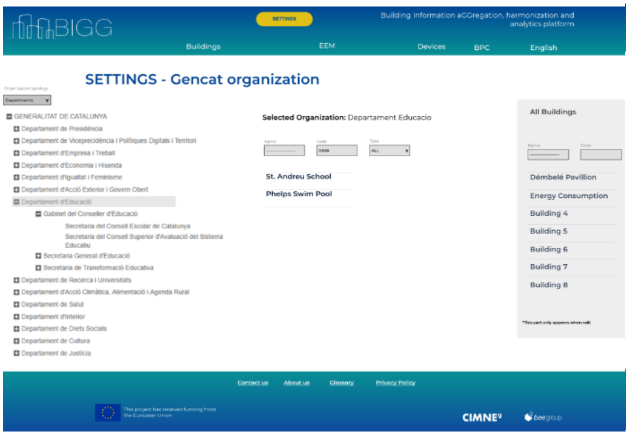
The basic functionalities implemented in this phase are: (a) Log-in, (b) Resource management, (c) Organization management; (d) User management; (e) List of buildings; (f) Building's form; (g) List of EEMs; (h) List of devices; (j) Individual device view.

Below are presented some representative mock-ups of the current user interface implementation.

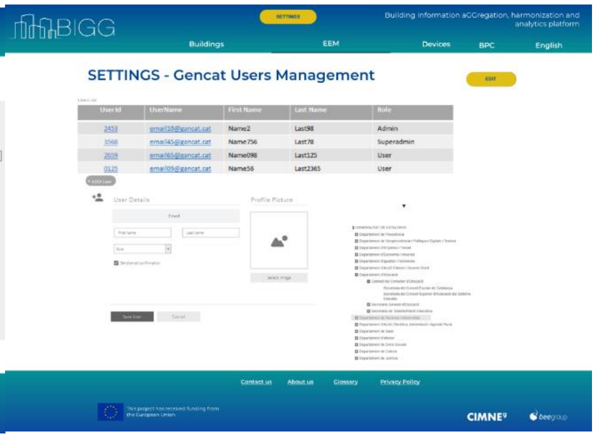
(a)



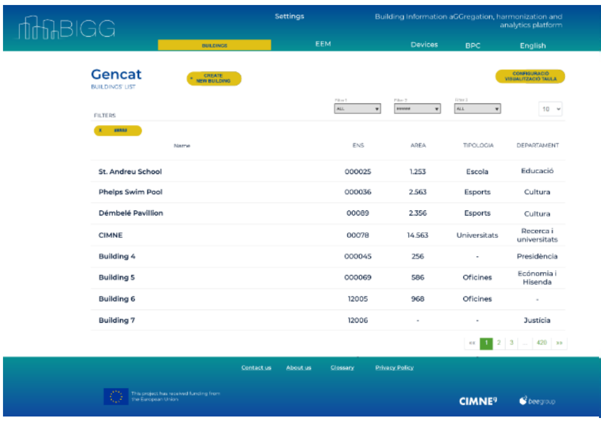
(b)



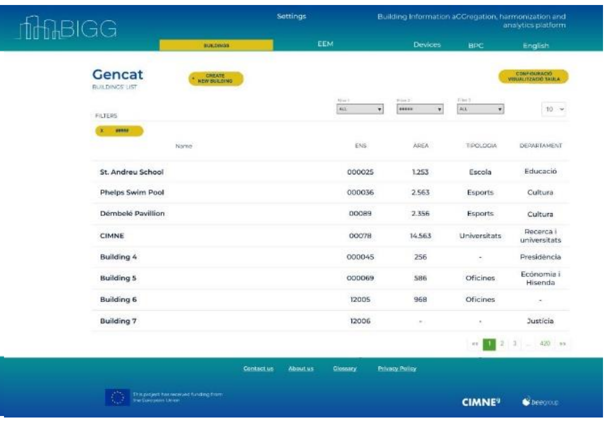
(c)



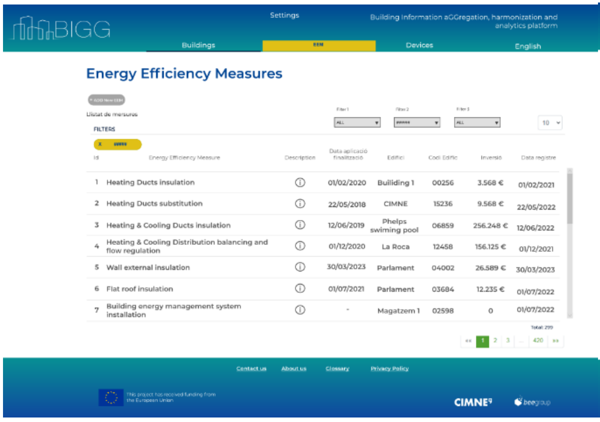
(d)

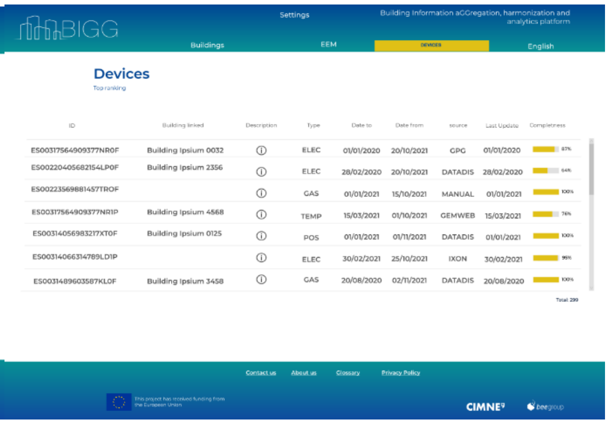


(e)



(f)





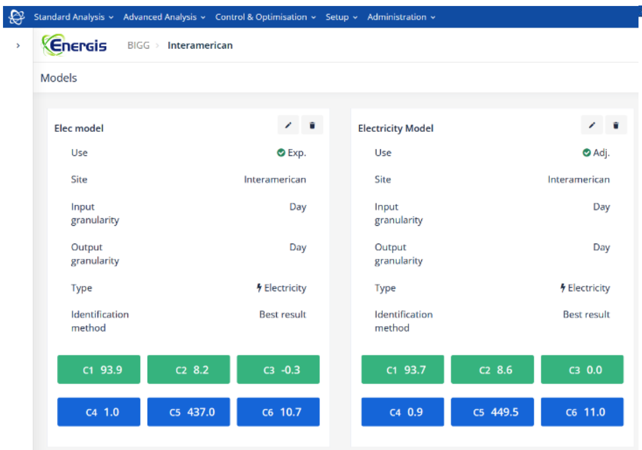
(g) (h)

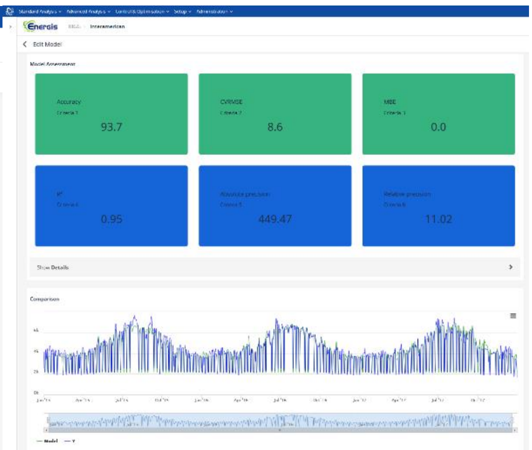
(j)

III.1.2.b. HELEXIA User Interfaces

For business cases 4 and 5 Cordia and Helexia will rely on two different user interfaces :

The first is the Energis platform which will be used to display time series including consumption data and models, weather data, indoor comfort data, energy savings from EPC projects, etc. This platform is used in the context of BC4 where data is being collected from the Cordia Connect platform (ex ENGIE connect) and used through the AI Toolbox to generate electricity consumption models and estimate savings in the context of an Energy Performance Contract management. The platform allows to track and manage identified models (figures a and b). It will be extended with the models from the AI toolbox.

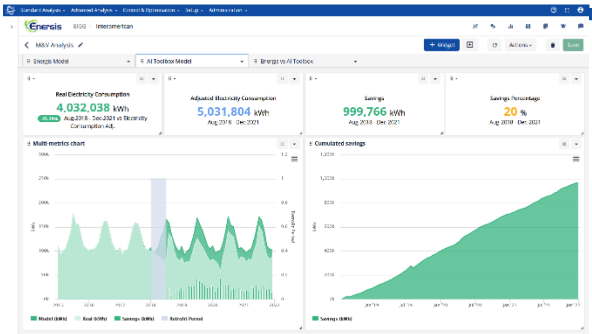




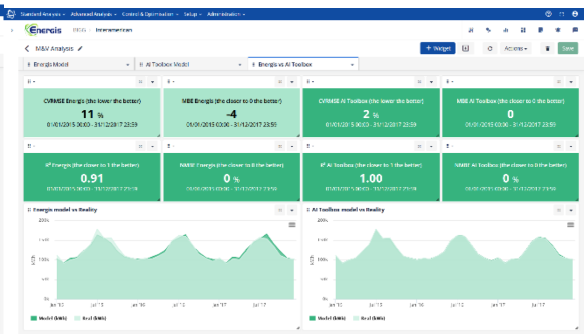
(a) (b)

From the identified models, the platform is used to compute the generated savings from the EPC by comparing the models with the actual building consumption after the retrofit period. These savings are then converted into different units and are compiled into a dashboard that can be followed in real time (figure c). The platform will also be used to periodically generate pdf reports about the EPC results which will be sent automatically to the EPC managers.

Since the pilot is also about demonstrating the improvements that can be expected from BIGG, the different models (business as usual models, built-in Energis models and the models generated from the AI toolbox) will be compared in a dashboard to highlight the key benefits from better models (figure d).

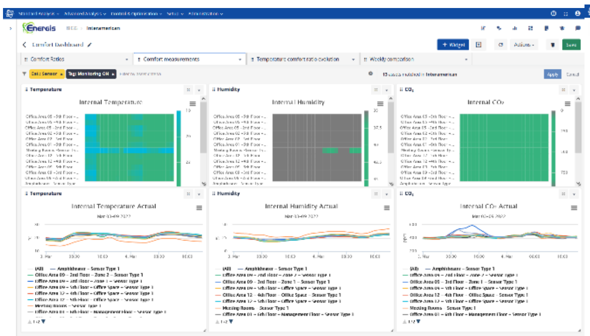


(c)

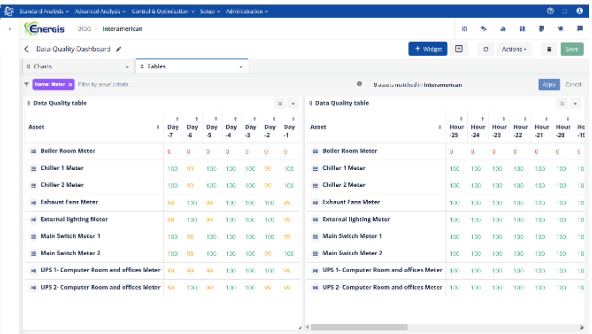


(d)

The Energis platform is also used in the context of BC5 to monitor comfort KPIs in the building resulting from the control actions that are put in place. For this purpose, a comfort dashboard is being used (figures e and f).

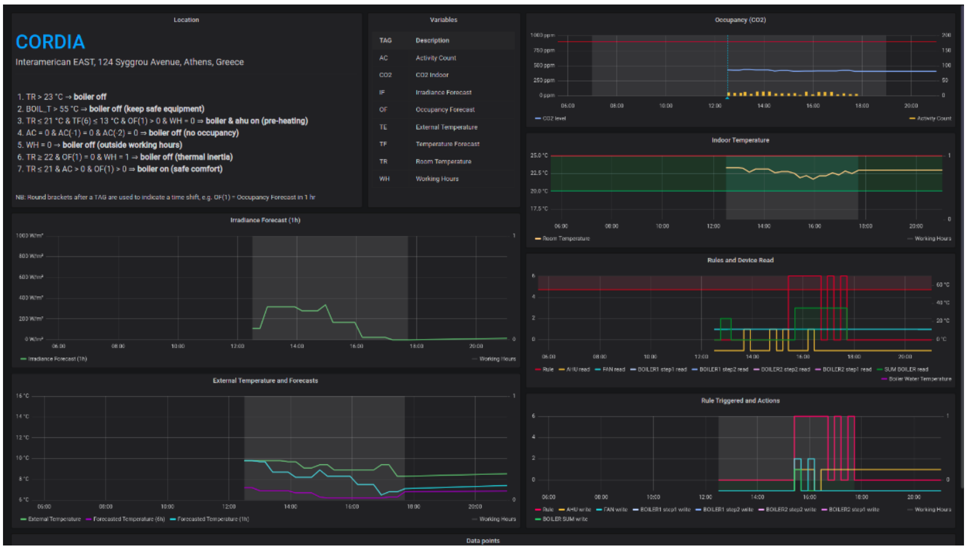


(e)



(f)

A second User Interface is 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 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.



(g)

This dashboard 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.

III.1.2.c. HERON User Interfaces

This use case provides the processes required to activate user flexibility for the Demand Response solution. The overal 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 will take place through an app which is currently under development. 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.

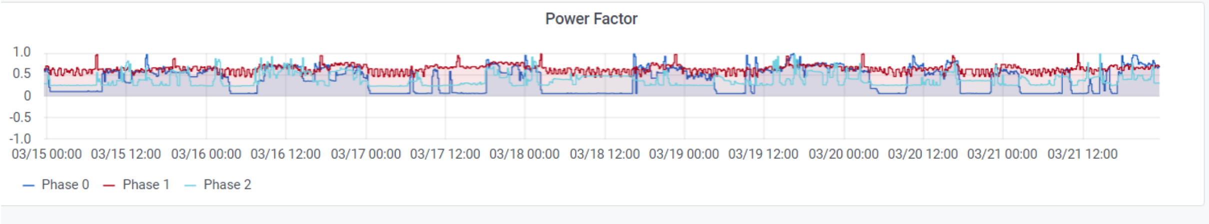


Figure 10 - Historical data for Power, Voltage, Energy and Power Factor per phase

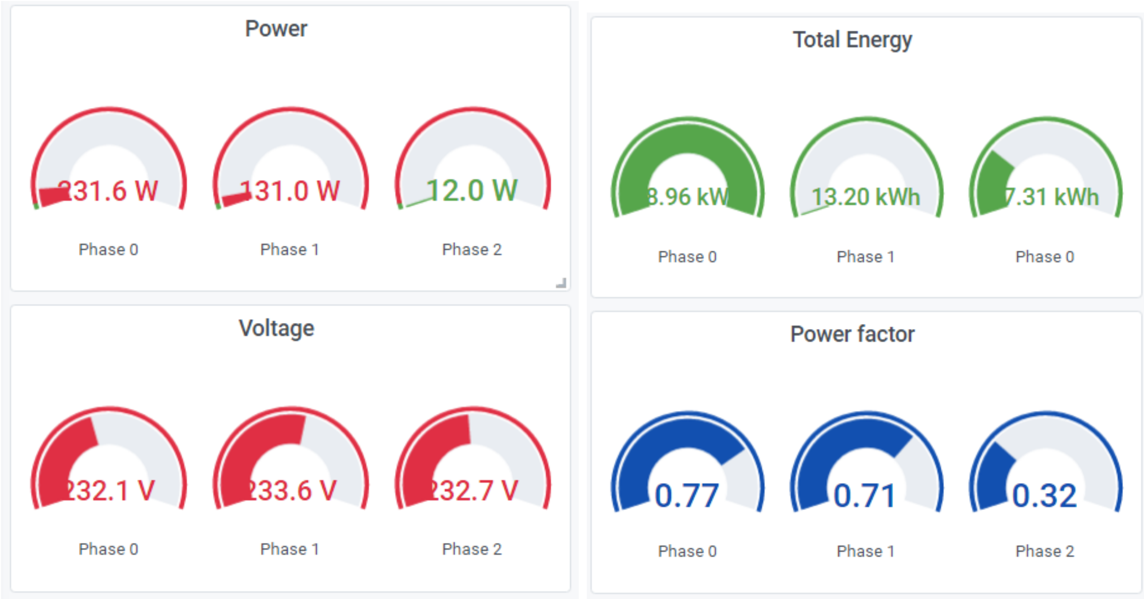


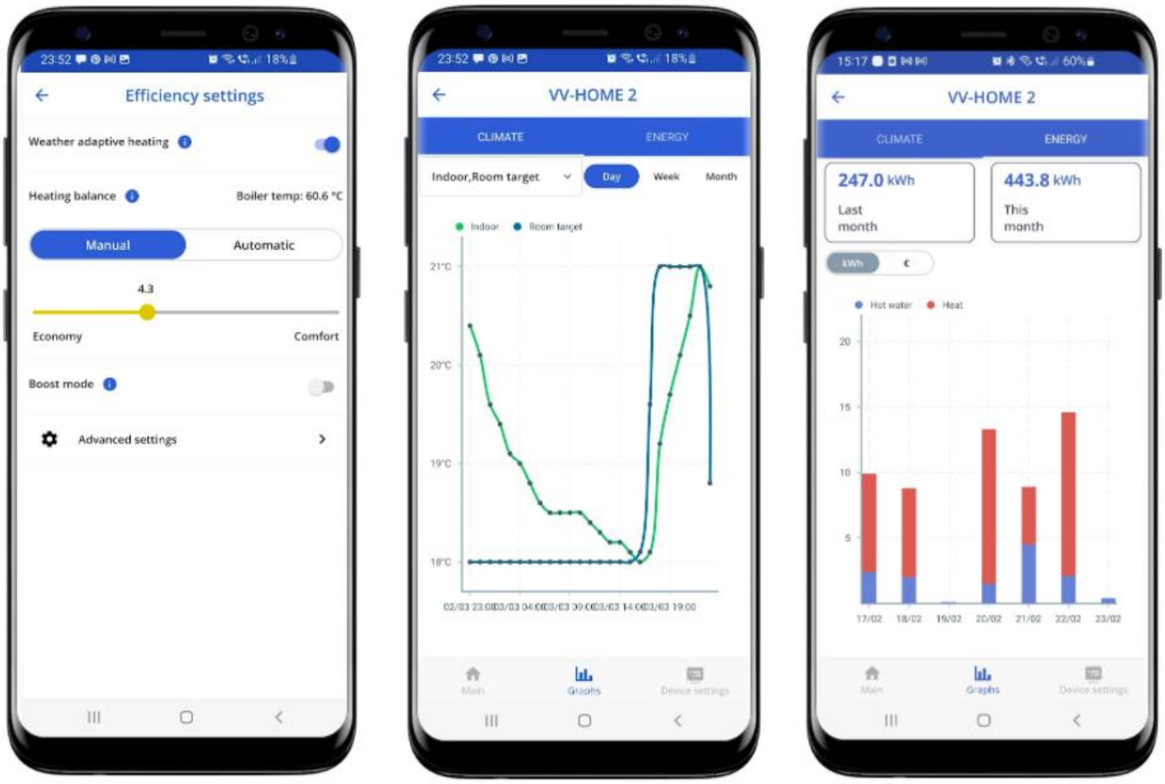
Figure 11 - Real-time data for Power, Voltage, Energy and Power Factor

III.1.2.d. 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.

End user Application

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 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 12 - DomX User Interfaces

Indicative screenshots of the extensions under development for the domX smartphone application employed by end consumers of UC15: (a) input of user comfort limits and opt-in / opt out to DR events (b) monitoring the room temperature deviation (c) monitoring the achieved gas consumption savings.

<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.

In the following Figure, several indicative mock-ups are illustrated to visualize the user interface under development that will implement various of the functionalities presented below:

- Real-time monitoring of portfolio demand (a)
- Real-time monitoring of natural gas balancing prices (a)
- Real-time monitoring of active boilers in map representation (b)
- Monitoring of buildings and available assets (c, d)
- Visualization of weather forecast
- Remote management of heating parameters (e)
- Visualization of forecasted demand and attained flexibility (f)

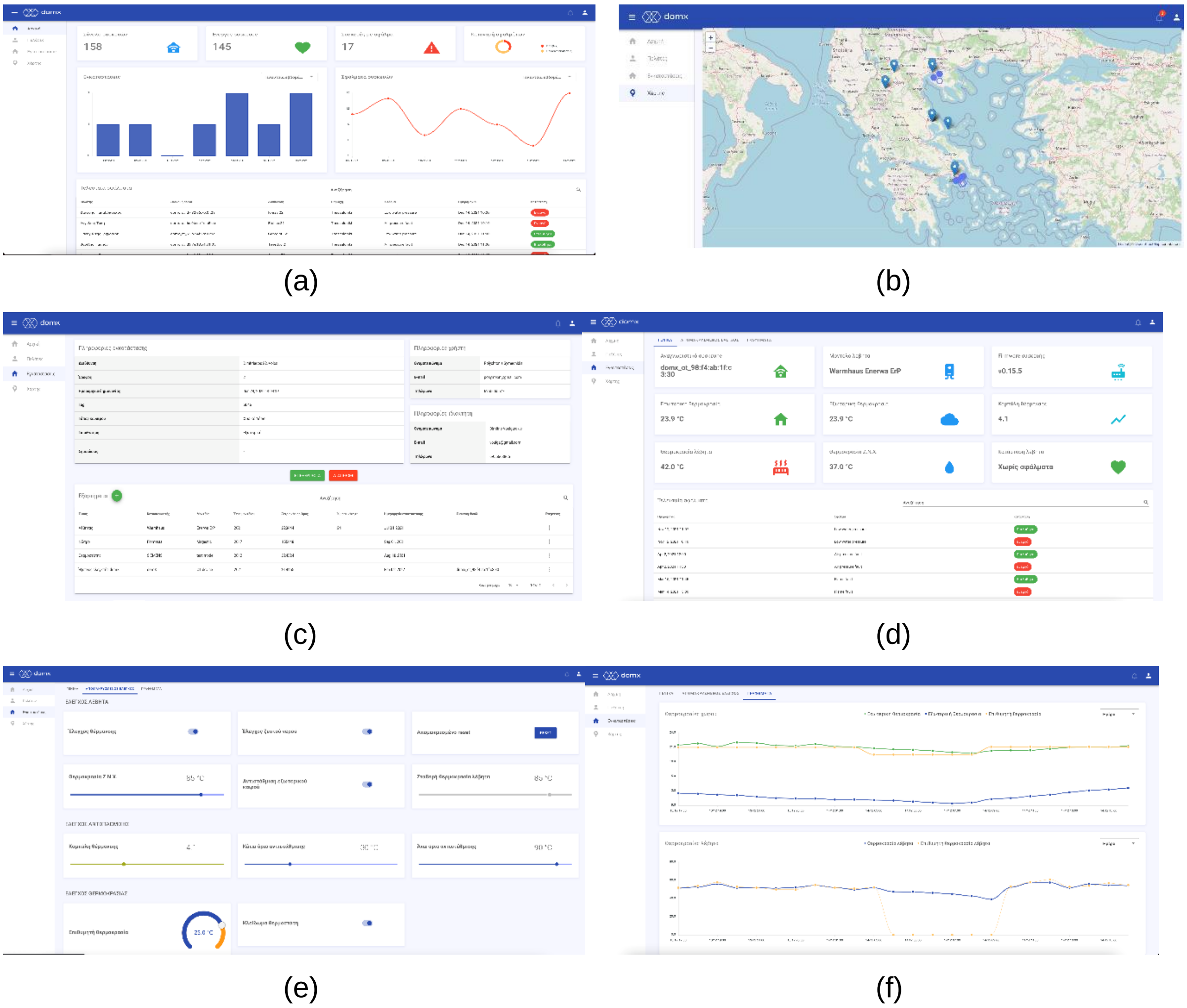


Figure 13 - Indicative screenshots of the Energy supplier's dashboard under development for UC15

IV. SECURITY

IV.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 minima as possible. But when we do need to process it, we will secure 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 IS27005[14] compatible 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 will group 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 will be 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 will 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 will be developed for each identified level of datasets in the data classification. For the determination of the security measures, we base ourselves on IS27002. The measures will be a grouping of organizational, governance, legal and technical measures. The defined set of minimal security measures will be 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 can help the use cases owners to implement these measures.
4. **Monitoring and assessment of applied security measures:** 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 requirements. Assesments will happen across all pilots by definition of KPI’s in terms of

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

organizational, governance, legal and technical levels, which will allow pilot evaluation and replicability.

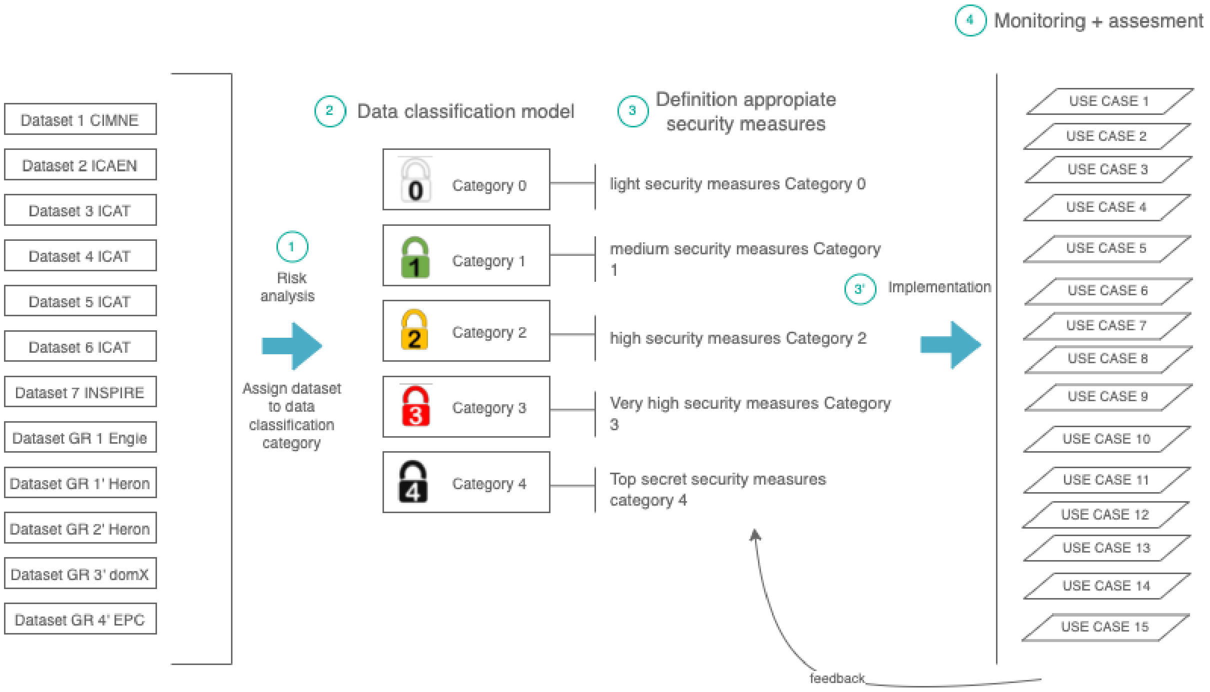


Figure 14 - Generic security approach

IV.1.1. Conclusion

We protect 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 develop a generic security approach. The data classification model, as well as the recommended minimal set of security measures for each classification level, will be packaged in the BIGG total solution to that future adopters of the BIGG framework can easily re-use this security approach.

IV.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.

BIGG security checklist

Question	Answer	SSRM ownership
re audit and assurance policies, procedures, and standards established, documented, approved, communicated, applied, evaluated, and maintained?	Yes	CSP owned
re 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	
re application security policies and procedures established, documented, approved, communicated, applied, evaluated, and maintained to guide appropriate planning, delivery, and support of the rganization's application security capabilities?	Yes	CSC owned
re strategies and capabilities established and implemented to deploy application code in a secure, standardized, and compliant manner?	Yes	CSC owned
re application security vulnerabilities remediated following defined processes?	Yes	CSC owned
re 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
re 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 15 - 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.

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
		Risk description				Risk calculation								
Nr	dataset ID	If	Then	Resource	Threat agent	Impact on reputation	Legal impact	Financial impact	Personal data pr	Average impact	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 rare	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 further 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 16 - Risk register

IV.3. Next Steps

Currently we are in the process of identifying and analysing all risks together with all partners. This will be finished by the end of M16. The further planning is as follows:[15]`

- Definition data classification model M17-M18
- Define minimal set of security measures: M19-M20
- Implementation security measures across all use cases: M21-M24

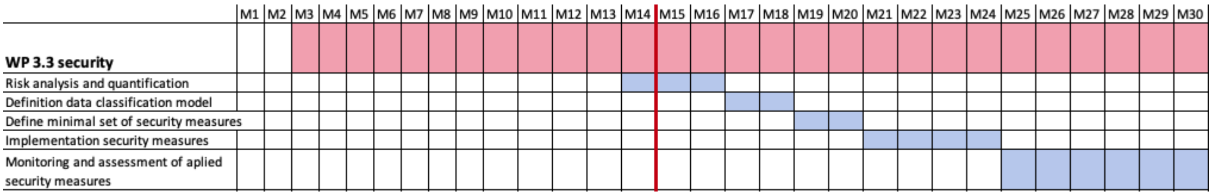


Figure 17 - Planning WP 3.3 Security layer development

[15] For use cases that will go-live before the generic minimal set of security measures is known an individual risk assessment will be done and individual security measures for the specific case will be defined and implemented.

V. CONCLUSION

V.1. Status

In this first year of 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 will serve as a validation for the architecture. Adjustments can and will be made when necessary.

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.

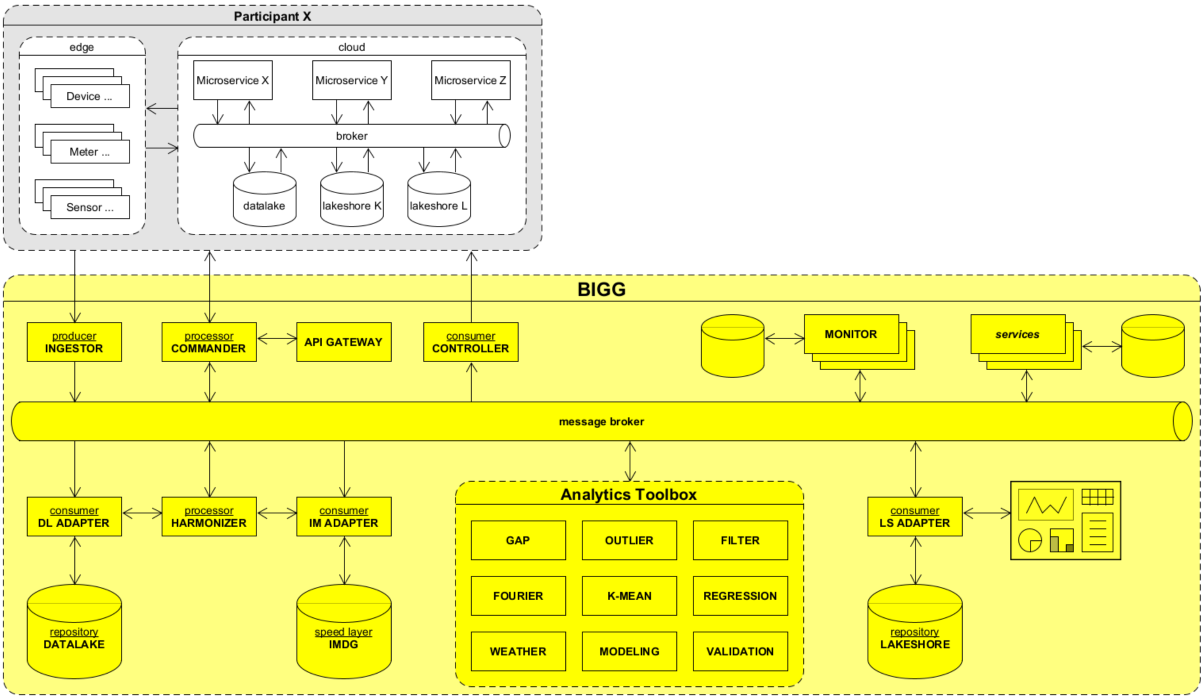


Figure 18 - BIGG Reference Architecture

All of these yellow blocks are the components that form BIGG. In Work Package 3 we focus on developing several of these blocks. The Communication Layer from Task 3.1 handles ingesting, harmonizing and processing 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.

In Work Package 3 the different business cases have created their own ingestor component, based on their needs. We have also started working on graphical user interfaces in the form of dashboards to monitor pilot sites and user applications per use case. The final task in this use case will provide guidelines on how to securely integrate the BIGG components in a solution.

V.2. Next Steps

In the coming period we will focus on aligning the business case architectures on the BIGG reference architecture. Meaning for example that we will take the business case specific versions of the ingestors and extract a generic ingestor, and ingestor skeleton from it. This will define a common way of ingesting data across all current and future business cases.