D2.6 – BEYOND Framework Architecture including functional, technical and communication specifications - a

Grant Agreement n°	957020
Project Acronym	**BEYOND**
Project Title	**A reference big data platform implementation and AI analytics toolkit toward innovative data sharing-driven energy service ecosystems for the building sector and beyond**
Starting Date	01/12/2020
Duration	36
EU Project Officer	Stavros Stamatoukos
Project Coordinator	UBITECH
Consortium Partners	VTT, FVH, CIRCE, Suite5, IGM, KONCAR, ARTELYS, MYTILINEOS, CUERVA, BELIT, URBENER, BEOELEK,
Project Website	beyond-h2020.eu
Cordis	https://cordis.europa.eu/project/id/957020

Deliverable No.	D2.6
Deliverable Title	**BEYOND Framework Architecture including functional, technical and communication specifications - a**
Work Package	WP2 - Use Cases, Business Requirements and Architecture Design
WP Leader	CIRCE
Due Date	31/09/2021
Actual Date of submission	25/10/2021
Version	V1.00
Status	Final
Dissemination Level	Public
Authors	Suite5
Reviewers	BELIT, CIRCE

Version	Modification(s)	Date	Author(s)
0.1	First ToC	26/03/2021	Suite5
0.2	Revised ToC	01/06/2021	Suite5
0.3	Revised Architecture Diagram	07/06/2021	Suite5
0.34	Sections 7.4, 7.5	14/7/2021	VTT
0.4	Sections 1,2,3,4,5, 7.6, 7.7	10/08/2021	Suite5, Ubitech
0.45	Section 7.9	25/8/201	KONCAR
0.55	Section 7.8	9/9/2021	BELLIT
0.6	Text Policing and Comments in existing Sections	9/9/2021	Suite5
0.7	Incorporation of further inputs in section 7 and addressing comments	20/9/2021	CIRCE, UBITECH, ARTELYS, FVH
0.8	Final Draft version ready for peer review	27/9/2021	Suite5
0.85	Final Contribution to sections 7.1, 7.2	13/10/2021	ARTELYS
0.90	Final Version addressing peer review comments	21/10/2021	Suite5
1.0	Final Version submitted to EC	25/10/2021	Suite5

EXECUTIVE SUMMARY

The deliverable at hand D2.6 “BEYOND Framework Architecture including functional, technical and communication specifications -a” aims at providing a detailed description of the BEYOND Platform architecture and its components, within the context of WP2 and more specifically with the initial work carried out of task T2.5 “Detailed architecture design, protocols and interfaces specifications for Big Data-enabled Energy Services”. In particular, this deliverable presents an overview of the conceptual platform architecture, a detailed description of its components and processes, its functional and non-functional requirements, as well as a detailed analysis of each end-user tool architecture and requirements.

The BEYOND platform aims to offer building and energy related data services, through the exploitation of its different data functionalities. Towards this end, the BEYOND framework’s design considers the following aspects:

- The provision of the overall BEYOND Platform conceptual architecture that consists of two core layers, i.e., the BEYOND Cloud-based Platform, and the BEYOND Private Infrastructure, both providing building and energy data-related functionalities either on the cloud or on-premise, respectively. Along with the overall conceptual architecture, the identification of the platform actors and the underlying core components and sub-components are described to provide a full architectural model of the BEYOND platform and highlight their role in the offered functionalities.
- The process description of the data-related functionalities, offered by the BEYOND Platform, that run either in background or involve user interaction. The processes views are classified into four core workflows, namely the Data Ingestion, the Data Exploration, the Data Trading, and the Data Analysis. All these processes are presented through Business Process Model and Notation (BPMN) diagrams, and the interactions with components and sub-components that are involved with each functionality are described accordingly.
- The extraction of functional and non-functional requirements of the BEYOND Platform, based on the intended platform functionalities and their relation to the different components and sub-components. An identification of each core component’s role and features in the BEYOND Platform and more specifically in the provided functionalities, is performed to extract the platform’s requirements.
- The detailed description of the end-user tools functionalities and architecture, as well as the description of the processes of the underlying elements of each end-user tool, such that the applications’ functional and non-functional requirements are defined appropriately.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

Abbreviation	Full Text
AI	Artificial Intelligence
APIs	Application Programming Interface
BEPO	Building Digital Twins Environment for Energy Performance Optimisation, Self-consumption Maximisation and Predictive Maintenance
BEYOND	A reference big data platform implementation and AI analytics toolkit toward innovative data sharing-driven energy service ecosystems for the building sector and beyond
BPMO	Building Portfolio Management Optimisation
CPOM	Cloud Platform Operations Manager
CIM	Common Information Model
Dx.y	Deliverable x.y
DAC	Data Analytics Container
DEM	Data Exploration Module
DGPIST	Distribution Grid Planning and Infrastructure Sizing Tool
DHCO	District Heating Network Planning and Infrastructure Sizing Tool
DIS	Data Ingestion Services
DLT	Distributed Ledger Technology
DoA	Description of Action
DTM	Data Trading Module
ETH	Ethereum
EPSRC	Energy Performance and Smart Readiness Certification
EPUL	Impact Assessment Tool for Energy Policy Making at Urban Level
FLEXopt	Flexibility-based VPP Configurator and DR Strategy Optimisation
IATEPM	Impact Assessment Tool for Energy Policy Making at Urban Level
IDAC	Isolated Data Analytics Containers
IPR	Intellectual Property Rights
PDIS	Private Data Ingestion Services
PDSL	Polyglot Data Storage Layer
PEASH	Personalised Energy Analytics for Guidance on Energy Performanc Optimisation and Human- Centric Control Automation
PIOM	Private Infrastructure Operations Manager
PPDSL	Private Polyglot Data Storage Layer
ROST	Renovation Optimisation Decision Support Tool
SEC	Security Components
Tx.y	Task x.y
WPx	Work Package x

1 Introduction

1.1 Scope of the document

This deliverable, D2.6 "BEYOND Framework Architecture including functional, technical and communication specifications - a" focuses on documenting the initial work carried out within the context of T2.5 "Detailed architecture design, protocols and interfaces specifications for Big Data-enabled Energy Services". This preliminary effort receives input from T2.1 "Definition of Business Scenarios, Use Cases and Elicitation of user & business requirements" in order to deliver an overview of the system architecture that describes the BEYOND Platform key components, and sub-components, the interfacing, and the different functionalities.

Within this framework, the major contributions of this deliverable are three: (a) to present and describe an overview of the conceptual architecture design of the BEYOND Platform, along with the different components, sub-components and the internal-external interfacing established between them; (b) to define the functional and technical specifications of the BEYOND Platform such that the dependencies among its different elements, and the resource constraints in terms of hardware and software, are clearly identified; and (c) to design the different components of the BEYOND Platform and to describe their intended functionalities along with their non-functional specifications and communication requirements.

Another major objective of this deliverable is to describe the different processes of the functionalities provided by the aforementioned components and sub-components. The processes' view describes explicitly the interactions of the involved components for all the main data-related functionalities that BEYOND Platform's users can follow for Data Ingestion, Data Exploration, Data Trading, and Data Analysis.

Finally, another aspect covered by this deliverable is the detailed architecture of the end-user tools provided by the BEYOND business developers. More specifically, this deliverable presents an overview of each end-user tool's architecture, defines their technical and non-technical requirements, and finally it describes the different components, their interconnection with the BEYOND Platform, and the process view of each end-user tool through sequence diagrams.

1.2 Relation to other tasks/deliverables

The deliverable at hand D2.6 "BEYOND Framework Architecture including functional, technical and communication specifications - a" receives inputs from the results obtained in T2.1 "Definition of Business Scenarios, Use Cases and Elicitation of user & business requirements" and reported in D2.1 "End-user & Business requirements analysis for big data-driven innovative energy services & ecosystems - a". In particular, D2.1 provides all the extracted end-user and business requirements for the big data-

driven services that are provided through the BEYOND Platform and draws certain specifications towards its design.

Moreover, D2.6 serves as a fundamental block of the architectural design of the BEYOND Platform, that is to be used for all the tasks under WP3 that are related to the development of the BEYOND Platform, i.e., T3.3 "Platform Backbone Infrastructure, On-Premise and Secure Experimentation Playground Data Containers and Core Services Development", T3.4 "Data Assets Security, Encryption and Privacy Mechanisms", and T3.5 "Platform and Services Bundles Continuous Integration and Open APIs Delivery".

An update of this deliverable, reporting on the efforts carried out for the design of the BEYOND Platform, will be provided in D2.7 "BEYOND Framework Architecture including functional, technical and communication specifications - b" due in M22 of the project's implementation.

1.3 Structure of the document

The remainder of the document is structured as follows:

- Chapter 2 provides a short description on the methodology followed for the design of the overall conceptual architecture of the BEYOND Platform based on the requirements elicitation of the different business scenarios and use-cases.
- Chapter 3 presents the key actors of the BEYOND Platform by identifying their role in the different platform processes.
- Chapter 4 presents an overview of the BEYOND Platform's conceptual architecture, along with the different aspects considered towards its design such as barriers and business and use cases requirements. Additionally, this chapter provides the description of the platform's core elements for both the BEYOND Cloud based Platform and the BEYOND Private Infrastructure.
- Chapter 5 provides a thorough description of the BEYOND Core Platform's architecture including the different components and their sub-components, as well as a description of the core platform's processes, ranging from data ingestion, data exploration, to trading and analysis.
- Chapter 6 presents the functional requirements as extracted from the different BEYOND Platform components, as well as the non-functional requirements of the overall BEYOND Platform.
- Chapter 7 explicitly describes the detailed architecture of each BEYOND end-user tool including its overall architecture, the components and sub-components that form the end-user tool, its functionalities and finally the extracted functional and technical requirements that are identified for each end-user tool.
- Chapter 8 concludes the deliverable and provides a short description of the next steps that are to be followed.

2 Methodology

Several project-relevant stakeholders, use cases, and business scenarios have been defined and described in D2.1 [2]to provide a better understanding of the building and energy sector, such that the business and system's requirements can be extracted appropriately. Such requirements along with the identification of the BEYOND system's actors (described in chapter 3), shall be used as the fundamental block for defining the design methodology of the BEYOND Framework Architecture.

The methodology adopted to design the BEYOND Framework Architecture is composed by five main steps that receive input from the Identification of Actors, and Business and User Requirements. As depicted in Figure 2- 1, the design methodology follows a sequential approach where the output or results of a certain step/block is utilised as input of the subsequent step/block.

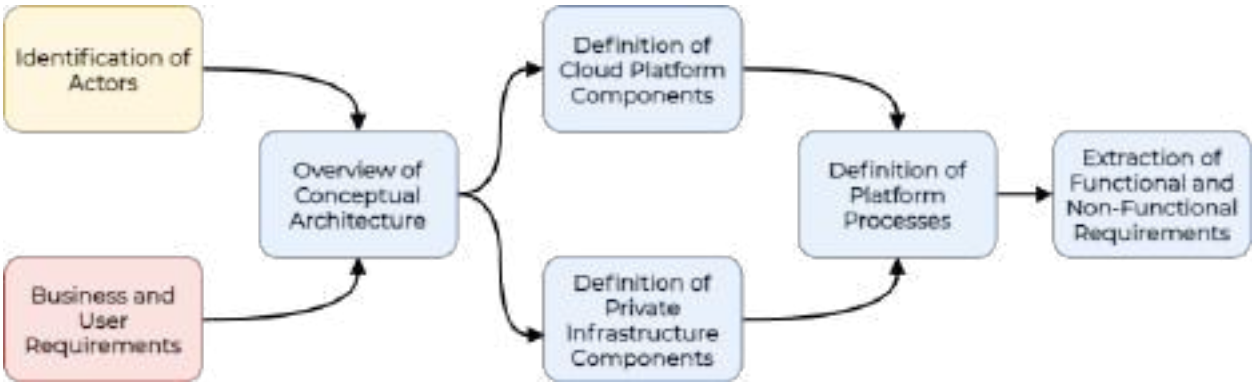

FIGURE 2- 1 BEYOND FRAMEWORK - DESIGN METHODOLOGY

The overview of the BEYOND Conceptual Architecture is defined mainly based on the business and user requirements as described in D2.1, while the role of the different actors, (described in chapter 3), are used to describe the different processes and elements of the BEYOND Framework.

The conceptual architecture of the BEYOND Framework is divided into two main layers, providing one layer for the cloud-based platform and one for the private infrastructure. These two layers are essential to provide a complete framework that can be used by the BEYOND Framework stakeholders to ensure that their data-related processes can be executed both on the cloud and on an organisation's private (on-premise) infrastructure. Based on the conceptual architecture, a more detailed description of the core components and subcomponents is provided, and the interactions among them are defined, in order draw the view of the different data-related processes provided through the different BEYOND Platform functionalities. Finally, based on the BEYOND layers, components and processes, a thorough list of the extracted functional and non-functional requirements is presented.

3 Identification of Actors

In this chapter, an identification of the actors intended to use the BEYOND Platform is performed and a grouping into six different classes of actors/users is presented in Figure 3- 1 below.

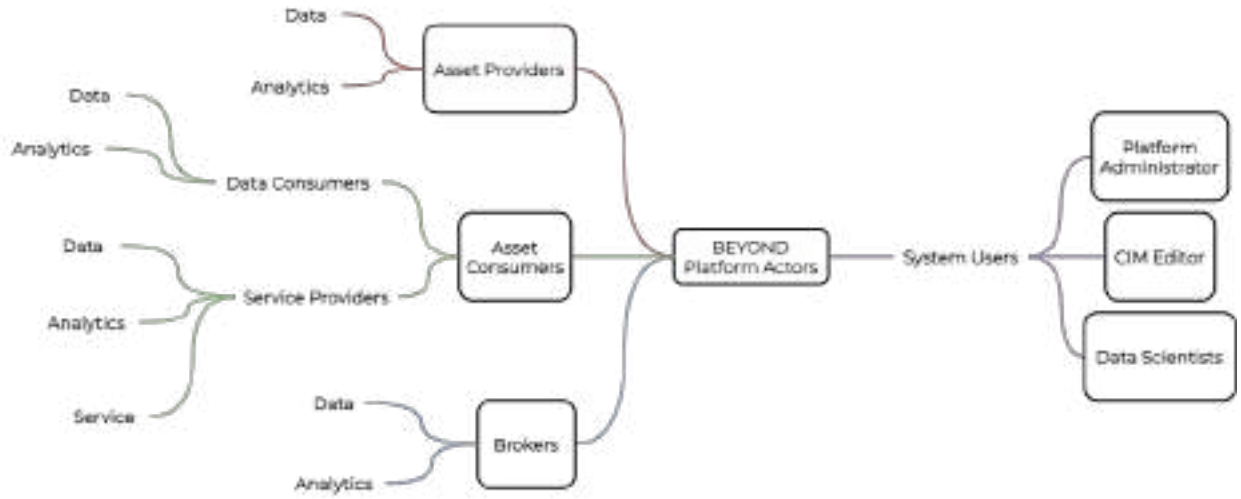

FIGURE 3- 1 BEYOND PLATFORM ACTORS

Platform Administrator: The Platform Administrator is responsible for monitoring the status of the BEYOND Platform and ensure its availability. Additionally, the Platform Administrator needs to perform any software or hardware-related upgrades, maintenance, and provide solutions to platform-related technical matters or failures that may occur.

CIM Editor: The CIM Editor is responsible for managing the BEYOND Common Information Model (CIM) and its continuous evolution, by monitoring its state, adding new concepts and fields, deprecating undesired concepts and fields, and reviewing requests from organisations for adding new concepts and fields to the CIM.

Data Scientists: Data Scientists are capable of extracting knowledge and valuable insights for a wide range of problems, from structured or unstructured data, by applying different manipulation functions and analytics algorithms over these data.

Asset Providers: Asset Providers are individuals or organisations that provide data, models, algorithms, and/or analytics results to be loaded to the BEYOND Platform in order to be shared and/or used by the different BEYOND Platform stakeholders.

Asset Consumers: Asset Consumers are individuals or organisations that have rights to access and use (depending on the data contracts) data, models, algorithms, and/or analytics results that are already loaded to the BEYOND Platform by Asset Providers.

Brokers: Brokers (or Information Brokers) are individuals or organisations responsible for collecting information and data, in order to sell or license them to third parties.

4 BEYOND High-Level Integrated Platform Architecture

4.1 Towards the BEYOND Architecture – Considerations from WP2

In this section, different considerations regarding the design phase of the BEYOND Architecture are identified and extracted mainly from the work performed in WP2. Such considerations include the platform development barriers, limitations and requirements that need to be considered during the design and development phase of the BEYOND Platform.

To ensure that the BEYOND Platform will be in alignment with the different stakeholders' needs, a list of business and end-user requirements was created and documented in D2.1, according to the defined BEYOND use cases and the results of an end-user survey conducted with the circulation of a relevant questionnaire. In brief, the business requirements include all necessary features for stakeholders ranging from data loading, mapping, cleaning, encryption, to data analysis, exploration, and trading. These platform functionalities reveal several aspects that need to be considered for the design of the BEYOND Architecture. Such aspects include the task execution orchestration, the common information model adaptation, as well as the data security and access mechanisms that need to be involved for the different data preparation, analysis, exploration, and trading features.

Another important aspect that needs to be considered during the design of the BEYOND Architecture, is the allocation of the different functionality elements (subcomponents) in their corresponding components. Although the components are designed to perform certain tasks, their subcomponents may have different conceptual design logic (see the different subcomponents labels in Figure 4-1), and they should be able to coordinate with other subcomponents even though they might belong to another component. The coordination between the platform's components should be effective to finalise the requested tasks while ensuring a secured information exchange among the components.

4.2 Conceptual architecture – overview

In this section, the overview of the BEYOND Platform including its core elements is described to state clearly the high-level role of each component placed in the BEYOND conceptual architecture. The BEYOND Platform consists of two conceptual layers where the BEYOND stakeholders can ingest, analyse, explore, and trade data either on cloud (i.e., BEYOND Cloud based Platform) or on the organisation's premises infrastructure (i.e., BEYOND Private Infrastructure)

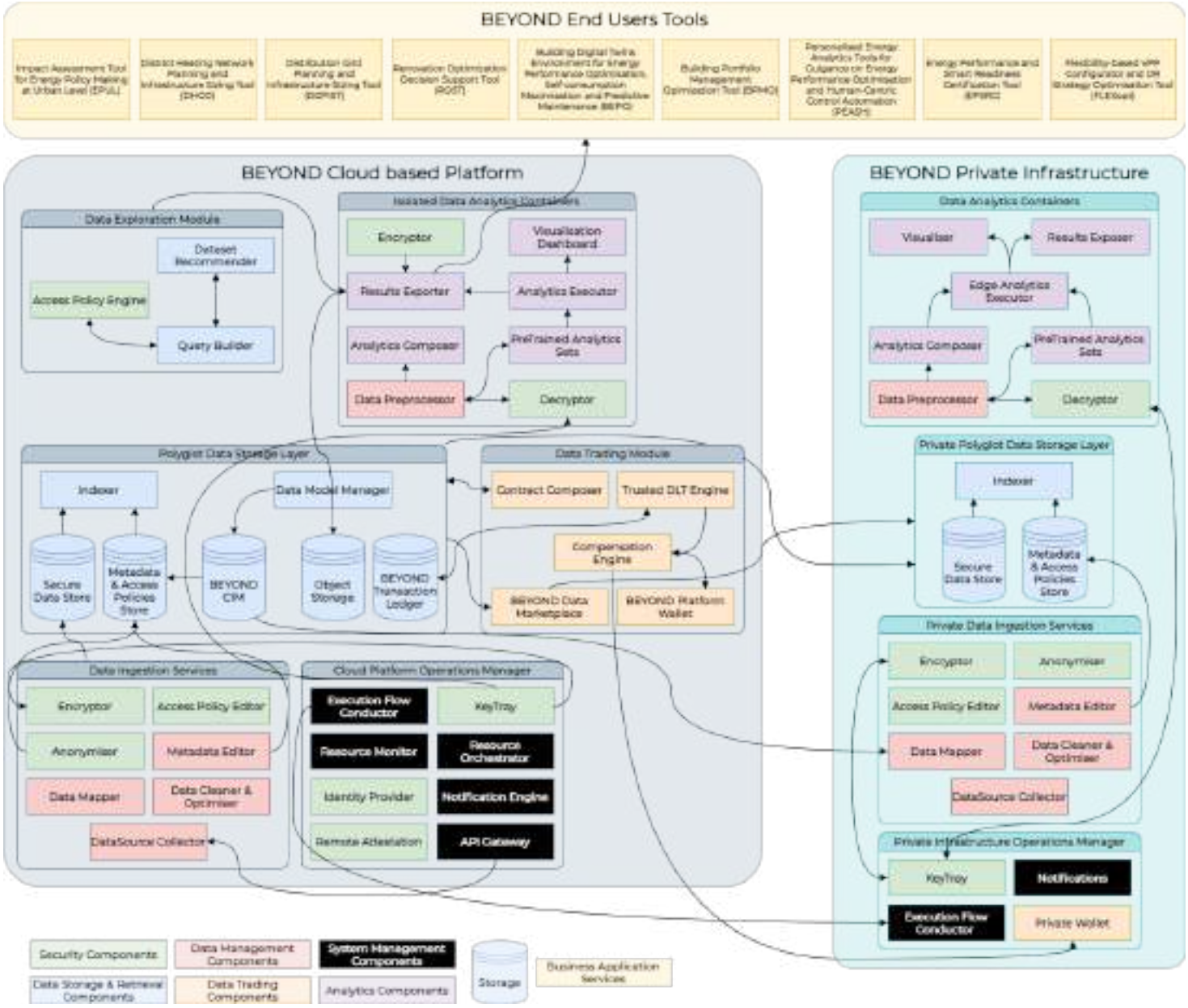

FIGURE 4-1: BEYOND CONCEPTUAL ARCHITECTURE

4.2.1 BEYOND Cloud based Platform

The BEYOND Cloud based Platform consists of five core components: (a) the Cloud Platform Operations Manager which is responsible to manage all the operations and resources within the cloud platform; (b) the Data Ingestion Services which is responsible to provide all the functionalities related to uploading data in the cloud platform; (c) the Polyglot Data Storage Layer storing temporary and permanent data, metadata, and other security information, needed for the different cloud platform functionalities; (d) the Isolated Data Analytics Containers which is responsible for designing and executing data analysis over the ingested data in the cloud platform; (d) the Data Exploration Module providing data search and exploration functionalities; and (e) the Data Trading Module which is responsible to prepare data contracts for sharing data among the different BEYOND users.

Data Ingestion Services

The Data Ingestion Services is mainly responsible for collecting data from external sources through the DataSource Collector component, and via batch data uploads, subscription to streaming data, or access to Application Programming Interfaces (API). In addition, this core element aims to manipulate the ingested data into an appropriate form for further analysis, by propagating the data into several sub-components that are responsible to clean data from outliers or missing values, map the data to a common information data model, add key details (metadata) to data, encrypt data denying access from unauthorised users, and protect individuals to be identified by anonymizing private or sensitive data. These functionalities are provided through data management components including: (a) DataSource Collector, (b) Data Cleaner & Optimiser, (c) Data Mapper, and (d) Metadata Editor; and security components including: (a) Access Policies Editor, (b) Encryptor, and (c) Anonymizer.

Polyglot Data Storage Layer

The Polyglot Data Storage Layer is responsible for storing all data ingested in the platform along with its metadata and access policies, the energy domain vocabularies and the common information model, the objects and models of algorithms and pipelines for data analytics, as well as all the information regarding data transactions. In particular, the Polyglot Data Storage Layer is comprised of five distinct storages namely: (a) the Secure Data Storage, (b) the Metadata & Access Policies Store, (c) the BEYOND CIM along with the Domain Vocabularies Store, (d) the Object Storage, and (e) the BEYOND Transaction Ledger. Additional data storage and retrieval components are included in the Polyglot Data Storage Layer, namely the Indexer and the Data Model Manager components that are responsible for coordinating the indexing of the ingested data and its associated metadata, and for the management of the common data model of the BEYOND Cloud based Platform respectively.

Data Exploration Module

The Data Exploration Module aims at providing the BEYOND Cloud based Platform with data browsing and exploration functionalities such that users that are interested in trading or acquiring particular data, to be able to view and request access on the potential data of interest. This module consists of the (a) DataSet Recommender, (b) the Query Builder, and (c) the Access Policy Engine, aiming to recommend data, based on predefined access policies, that might be potentially useful for certain platform's users.

Isolated Data Analytics Containers

The Isolated Data Analytics Containers have a fundamental role within the BEYOND Cloud based Platform since they aim to provide all functionalities for designing data analytics processes, based on machine learning and deep learning algorithms, towards the reporting or visualisation of the corresponding results. In particular, these containers include: (a) the Data Preprocessor, (b) the Analytics Composer, (c) the PreTrained Analytics Sets, (d) the Analytics Executor, (e) the Visualisation Dashboard, and (f) the Results Exposer. Additionally, to support encryption on the results and data access on the encrypted input data, the Isolated Data Analytics Containers include Decryptor and Encryptor components, respectively.

Data Trading Module

The Data Trading Module aims at providing trustful and secure trading services of data between the users of the BEYOND Cloud based Platform. This module consists of five core data trading components: (a) the BEYOND Data Markeplace, (b) the Contract Composer, (c) the Trusted DLT Engine, (d) the Compensation Engine, and (e) the BEYOND Platform wallet, providing all the functionalities needed under the context of data sharing and trading.

Cloud Platform Operations Manager

The Cloud Platform Operations Manager is mainly responsible for the management and coordination of the various platform's components, the orchestration of the underlying resources, and the management of appropriate notifications, while ensuring its secure execution and operation. In particular, the Cloud Platform Operations Manager consists of five core system management components, namely: (a) the Execution Flow Conductor, (b) the Resource Monitor, (c) the Resource Orchestrator, (d) the Notification Engine, and (e) the API Gateway, as well as three distinct security components, namely: (a) the KeyTray, (b) the Identity Provider, and (c) the Remote Attestation.

4.2.2 BEYOND Private Infrastructure

The BEYOND Private Infrastructure consists of four core components: (a) the Private Data Ingestion Services which is responsible to provide all the functionalities related

to uploading data in the private infrastructure; (b) the Private Polyglot Data Storage Layer storing temporary and permanent data, metadata, and other security information, needed for the different private infrastructure functionalities; (c) the Data Analytics Containers responsible for designing and executing data analysis over the ingested data in the private infrastructure; and (d) the Private Infrastructure Operations Manager which is responsible for managing the different operations within the private infrastructure, as well as for communicating with the Cloud Operations Manager to coordinate the execution of the various BEYOND functionalities.

Private Data Ingestion Services

The Private Data Ingestion Services component is responsible for collecting and preparing data as in the case of the Data Ingestion Services of the BEYOND Cloud based Platform, in a private and secure environment located in the organisation's infrastructure. Although the various processes offered by the components of the Private Data Ingestion Services are almost identical to the ones offered by the components of the Data Ingestion Services of the BEYOND Cloud based Platform, the configuration and execution of the data ingestion processes are performed locally in the organisation's private infrastructure to ensure increased privacy.

Private Polyglot Data Storage Layer

The Private Polyglot Data Storage Layer is responsible to store the ingested data in organisation's local premises for increased security. Thus, it only consists of two storage components, namely the Secure Data Store and the Metadata & Access Policies Engine Store, while an Indexer component is used to facilitate the retrieval of stored data in an efficient way. For additional storage capabilities, the Private Polyglot Data Storage communicates with the Cloud based Platform to invoke the execution of additional functionalities.

Data Analytics Containers

The Data Analytics Containers located in the BEYOND Private Infrastructure on organisation's premises, provides all functionalities needed for designing and executing data analytics processes, based on machine learning and deep learning algorithms, towards the reporting or visualisation of the corresponding results. Like the Isolated Data Analytics Containers, the Data Analytics Containers includes: (a) the Data Preprocessor, (b) the Edge Analytics Composer, (c) the PreTrained Analytics Sets, (d) the Analytics Executor, (e) the Visualiser, and (f) the Results Exposer. Although Decryption functionalities are needed for accessing the ingested data that are to be further processed, the Data Analytics Containers component does not include an Encryptor component for the corresponding results since these remain in the organisation's local premises.

Private Infrastructure Operations Manager

The Private Infrastructure Operations Manager receives execution flow commands from the Cloud Operations Manager to coordinate the execution of the various BEYOND Private Infrastructure components with the Flow Conductor component. Additionally, it consists of a KeyTray component for providing the encryption keys to the Private Data Ingestion Services, and a Private Wallet that communicates with the Compensation Engine of the BEYOND Cloud based Platform, allowing secure management of the ledger account and data contracts. A Notification component is also present to provide the BEYOND Private Infrastructure with appropriate messages regarding events, tasks, and their corresponding progress, relevant to the users' preferences.

5 BEYOND Core Platform – Detailed Architecture

5.1 Components View

Overall, the BEYOND Cloud based Platform is comprised of six components groups; each of them designed with a specific set of functionalities and purpose and which are namely: a) the *Data Ingestion Services* providing the means to collect data to the platform; b) the *Polyglot Data Storage Layer* which is comprised of many different data storing technologies necessary to facilitate the different components; c) the *Data Exploration Module* that is actually working with the different data stores to search for data assets and provide access to the data assets available in the BEYOND platform; d) the *Isolated Data Analytics Containers,* where users can set up and run analytics in a security and privacy preserving manner; e) the Data Trading Module which is in charge of acquiring the data assets out of the stores and allowing stakeholders to trade data through multiparty smart contracts being supported by blockchain technology and finally f) the *Cloud Platform Operational Manager,* including a bundle of components that have to do with the efficient operation of the platform such as identity provider, users' authentication, monitoring of the resources and orchestrating the operation and exchange of messages, data and control between the BEYOND Cloud based Platform and the BEYOND Private Infrastructure.

Having provided an overview of the BEYOND Cloud based Platform the following sections provide a detail view of the aforementioned components groups, detailing their relevant subcomponents, the lists of functional requirements as extracted from the components and the technology stack on which they are built.

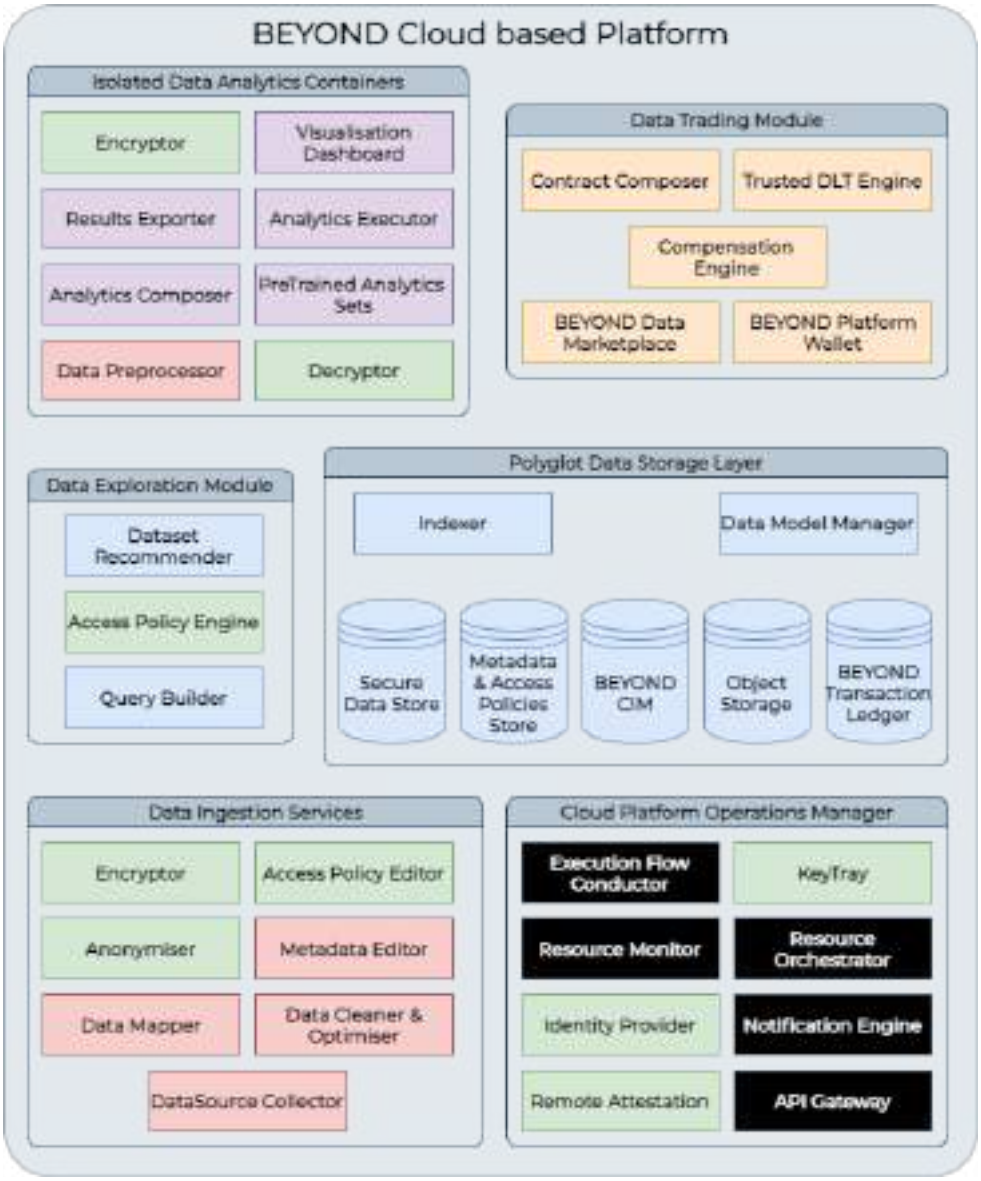

FIGURE 5-1: BEYOND CLOUD BASED PLATFORM: CONCEPTUAL ARCHITECTURE

5.1.1 Cloud Based Platform – Detailed Architecture

5.1.1.1 Data Ingestion Services

Data Ingestion Services allow BEYOND stakeholders, acting as data owners/providers to collect, load, and prepare building-relevant data from different sources for further analysis, and finally make them available to the BEYOND Cloud based Platform. The various components included in the Data Ingestion Services are classified into two different categories based on their mission of handling data. In particular, Data Ingestion Services component is depicted in Figure 5-2, and it is comprised of Data Management Components including: (a) the DataSource Collector, (b) the Data Cleaner & Optimiser, (c) the Data Mapper, and (d) the Metadata Editor; and Security

Components including: (a) the Access Policies Editor, (b) the Encryptor, and (c) the Anonymizer.

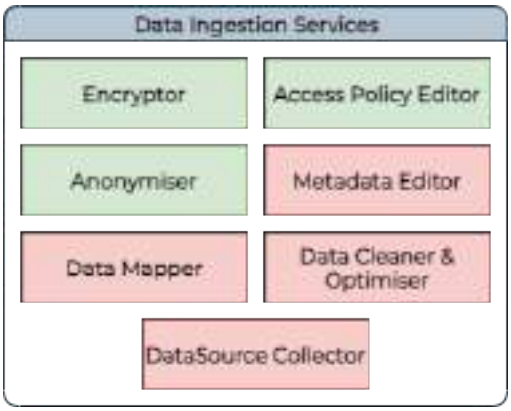

FIGURE 5-2: DATA INGESTION SERVICES: COMPONENT OVERVIEW

Building-relevant data can be loaded on the platform by configuring and executing the DataSource Collector component which is responsible to collect the data through one of the following three data ingestion methods: (a) the Batch File Upload, (b) the API Request, or (c) the Subscription to Streaming Data. Thus, an appropriate configuration on the data ingestion method should be defined accordingly to ensure that all prerequisite actions are taken into consideration prior to the actual data ingestion execution. Upon loading the data on the platform, data owners/providers should configure the Data Cleaner & Optimiser component that is responsible for cleaning the data and ensuring it is correct, consistent and usable. The execution of a configured cleaning process involves the identification of corrupted or missing data, and its replacement with appropriate values such that it can be further processed for its intended use. As a next step, Data Ingestion Services allow mapping of the ingested data to the BEYOND CIM by configuring and executing the Data Mapper Component. Such a process ensures that a common data model, that is regularly maintained and updated based on different existing standards, is established to preserve a common understanding among the BEYOND stakeholders. Along with the mapping of the ingested data, the Metadata Editor component aims to provide appropriate and explicit description about the data to facilitate its use, mainly in the data exploration process within the BEYOND Cloud based Platform.

Subsequently, once all the data management processes within the Data Ingestion Services have been configured, data owners/providers should define certain data access policies on their ingested data. This configuration is handled (and stored in the Metadata & Access Policies Store of the Polyglot Data Storage Layer) by the Access Policies Editor which is responsible for determining appropriate access policies on the data based on the data owners/providers' needs. The data access policies range from public, to restricted, and internal, denoting if the data is intended to be available to all, to other organisations who have signed data access contracts, or only to the data owner/provider's organisation, respectively. An additional Security Component included in the Data Ingestion Services is the Encryptor. This component is

responsible for ensuring that the information describing a dataset, or a part of it, should be encoded, denying the access to unauthorised users. Encryptor allows data owners/providers to select the parts of data that are to be secured, while a unique key for decrypting the data when needed is generated and obtained automatically from the KeyTray (included in the Cloud Platform Operations Manager). For sensitive and private information that could potentially point to a particular individual, data anonymisation processes should be defined as well. In fact, the Data Ingestion Services provides the Anonymizer component which is responsible to protect sensitive data by apply different methofd, such as removing any undesirable identifier that could reveal private information, by defining certain anonymisation rules on the data under consideration.

Functional Requirements

All data management components included in the Data Ingestion Services (i.e., DataSource Collector, Data Cleaner & Optimiser, Data Mapper, and Metadata Editor) provide to the data providers/owners various functionalities for configuring and executing data management functions such as ingestion, cleaning, and mapping. These functionalities are extracted based on their direct relation with the aforementioned data management functions, as well as on the end-user and business requirements as reported in D2.1 "End-user & Business requirements analysis for big data-driven innovative energy services & ecosystems - a". In particular Table 5- 1 presents the functional requirements extracted regarding the Data Ingestion Services.

TABLE 5- 1 DATA INGESTION SERVICES – FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
DIS_01	The BEYOND solution shall allow data providers to upload their data to the BEYOND Cloud based Platform in different file formats (e.g., csv, json, etc).	UC_D_3.3
DIS_02	The BEYOND solution shall allow data providers to load streaming data to the BEYOND Cloud based Platform.	UC_D_3.3
DIS_03*	The BEYOND solution shall support the temporal configuration of the data loading services (e.g., schedule data loading).	UC_D_3.3
DIS_04	The BEYOND solution shall support the execution of scheduled data uploading in the BEYOND Cloud based Platform.	UC_D_3.3
DIS_05*	The BEYOND solution shall support the execution of scheduled and periodic data loading.	UC_D_3.3

*** Denotes that a particular functional requirement is also extracted in the Private Data Ingestion Services component.

DIS_06	The BEYOND solution shall support the execution of scheduled collection of streaming data.	UC_D_3.3
DIS_07	The BEYOND solution shall support the execution of scheduled collection of data from REST API.	UC_D_3.3
DIS_08*	The BEYOND solution shall validate that the ingested data are of appropriate format, range, and content.	UC_D_3.1
DIS_09*	The BEYOND solution shall allow data providers to set rules for handling incomplete or corrupted entries, and outliers in their ingested data.	UC_D_3.1
DIS_10*	The BEYOND solution shall allow data providers to set rules for filling empty entries (possibly detected) in their ingested data.	UC_D_3.1
DIS_11*	The BEYOND solution shall support the fusion of different data formats into homogenous data format.	UC_D_3.1
DIS_12*	The BEYOND solution shall ensure that a mapping between data providers' ingested data to the common information model is established.	UC_D_3.1
DIS_13*	The BEYOND solution shall allow data providers edit and update the defined mapping.	UC_D_3.1
DIS_14*	The BEYOND solution shall allow data providers to select transformation rules (e.g., measurement units, measurement scale, time zones, etc) to be applied on their ingested data, and according to the common information model.	UC_D_3.1
DIS_15*	The BEYOND solution shall enable semantic interlinking of different entities within the platform.	UC_D_3.1
DIS_16*	The BEYOND solution shall ensure that the quality of the ingested data are of high quality by combining them with other data on a semantic and payload level.	UC_D_3.2
DIS_17*	The BEYOND solution shall support the inclusion of various metadata on the ingested data, to improve asset discoverability, and reveal relationships that might exist with other assets.	UC_D_3.2
DIS_18*	The BEYOND solution shall enable the fusion of different assets into a composite asset, to create new data profiles that carry more valuable information.	UC_D_3.2
DIS_19*	The BEYOND solution shall ensure that data transformations for measurement units are in accordance with the BEYOND CIM.	UC_D_3.1

Technology Stack

The Data Ingestion Services provide all the functionalities and services, extracted from the functional requirements as described in the previous section, through various state-of-the-art technologies that are presented below:

Back-end layer:

a) Nest (NodeJS)[1] web framework for delivering efficient, reliable, and scalable server-side applications,
b) Flask[2] micro web framework,
c) Pandas[3] for handling, cleaning, and anonymizing data,
d) Kafka[4] distributed stream-processing software platform as the publication-subscription mechanism for streaming data collection,
e) RabbitMQ[5] message broker system for exchanging feedback messages, and keys to the Cloud Platform Operations Manager component.

Front-end layer:

a) VueJS[6],
b) TailwindCSS[7] for custom front-end design.

Data storage layer:

a) PostgreSQL[8] for storing the configurations of data ingestion, cleaning, anonymisation, encryption,
b) MinIO[9] for the temporarily storage of the ingested, cleaned, anonymised, encrypted, data and intermediate data management,
c) Vault[10] for securing sensitive and secret parameters.

5.1.1.2 Polyglot Data Storage Layer

Overall, the Polyglot Data Storage Layer (PDSL) represents the central secure storage layer of the BEYOND Cloud based platform. Its main responsibility is the persistent storage of all the various data assets (such as datasets, analytics models, analytics results and reports, ingestions job-specific data, etc.) along with their associated metadata and access policies, in a secure and reliable manner; also enabling

[1] https://nestjs.com/
[2] https://flask.palletsprojects.com/en/2.0.x/
[3] https://pandas.pydata.org/
[4] https://kafka.apache.org/
[5] https://www.rabbitmq.com/
[6] https://vuejs.org/
[7] https://tailwindcss.com/
[8] https://www.postgresql.org/
[9] https://min.io/
[10] https://www.vaultproject.io/

encrypted storage of all relevant confidential data and credentials (e.g., tokens, API keys, usernames and passwords). Moreover, the PDSL enables storage of all administrative information necessary for the efficient operation and usage of the BEYOND Cloud based Platform, the BEYOND Private Infrastructure and the BEYOND End Users Tools. The PDSL has a two-fold scope; it securely stores all data assets that are ingested and/or generated in the BEYOND platform, while also makes such data assets retrievable (i.e., available) to the rest of the BEYOND Platform modules and applications. As depicted in Figure 4-1, the Polyglot Data Storage Layer is compiled by both Storage components and Data Storage and Retrieval components, whose functionalities and communication interfaces are described as follows.

In respect to the Storage components, once the data assets have been ingested and properly processed through the Data Ingestion Services, these are permanently stored in the *Secure Data Store*, facilitating optimised management of the stored data assets. The associated metadata and enforced access policies, along with the links to the related data assets are stored in the *Metadata & Access Policies Store*. The *BEYOND CIM and Domain Vocabularies* container (being the output of T3.2) is responsible for storing the various versions of the BEYOND CIM along with the related domain-specific vocabularies, thus enabling though communication with the Data Mapper (included in the Data Ingestion Services, see section 5.1.1.1) the mapping of the different information entities to the BEYOND CIM concepts.
The *Object Storage* container is utilised to store the datasets being the outcomes of a data analytics job (such as analytics models, analytics results along with their reports) in order to be available to the other BEYOND platform's modules and end-users tools. Lastly the *BEYOND Transaction Ledger* is responsible for storing the various data asset contract-related information, while securing the privacy of the involved stakeholders and providing full traceability.

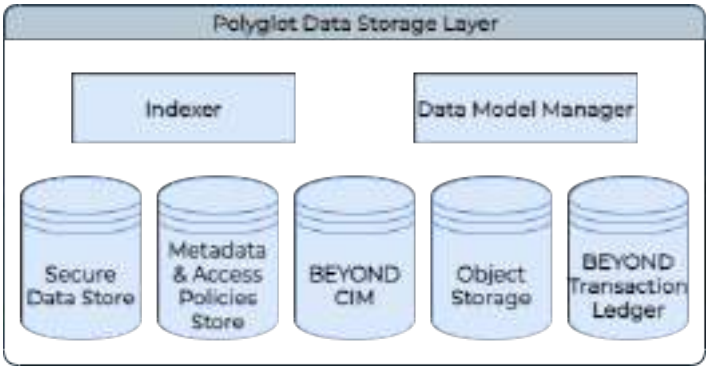

FIGURE 4- 1 POLYGLOT DATA STORAGE LAYER: COMPONENT OVERVIEW

Accordingly, the Data Storage and Retrieval components included in the Polyglot Data Storage Layer include a) the *Indexer*, responsible for managing the indexing of the stored data assets (and their metadata) so as to be easily searchable, while respecting the access policies set by their respective data providers and b) the *Data*

Model Manager, responsible for the management of the BEYOND CIM and the Domain Vocabularies.

As mentioned previously the Polyglot Data Storage Layer provides a variety of data storage and data retrieval & indexing components, mainly related to the BEYOND Cloud based platform's storage needs. Such storage needs are directly linked with the end-users' and business requirements as reported in D2.1 [2] and which are translated in Table 5- 2 into a set of functional requirements that shall be met by the Polyglot Data Storage Layer.

TABLE 5- 2 POLYGLOT DATA STORAGE LAYER - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
PDSL_1	The BEYOND solution shall provide a consolidated storage component for the secure storage of all data assets ingested by the platform.	UC_D_3.3
PDSL_2*	The BEYOND solution shall enable data providers to store their data assets in a trusted data container, enforcing data security and privacy- ensuring mechanisms.	UC_D_3.4
PDSL_3*	The BEYOND solution shall enable storage of the data assets' metadata in a trusted data container.	UC_D_3.4
PDSL_4*	The BEYOND solution shall enable the storage of the access policies defined by the data providers on their assets.	UC_D_3.4
PDSL_5:	The BEYOND solution shall enable storage of the data analytics job related information's (e.g., trained algorithms, analytics models and results) in a trusted data container.	UC_D_3.4
PDSL_6*	The BEYOND solution shall enable the storage of all data ingestion job-specific configurations along with the transitional data files, enabling data providers if needed, to pause of a data ingestion job and undertake updates or resume it at a later stage.	UC_D_3.3
PDSL_7*	The BEYOND solution shall provide a secure storage container for saving all sensitive data and credentials, while enforcing controlled access over the different sensitive parameters, such as tokens, API keys, usernames, passwords, etc.	UC_D_3.4
PDSL_8	The BEYOND solution shall provide a secure storage container for saving all the various versions of the BEYOND CIM, inclusive of the model's concepts, fields and their relations.	UC_D_3.4

*** Denotes that a particular functional requirement is also elicited for the Private Polyglot Data Storage Layer.

PDSL_9	The BEYOND solution shall provide a secure storage container for saving all the various versions of the domain vocabularies.	UC_D_3.4
PDSL_10	The BEYOND solution shall provide a dedicated distributed ledger for the secure storage of all contract-related data, while ensuring the privacy of the involved stakeholders.	UC_D_3.4

Technology Stack

Regarding the technology stack of the Polyglot Data Storage Layer, its Data Storage layer is built on state-of-the art storage and indexing technologies utilising:

a) PostgreSQL[8] as the BEYOND Cloud based platform's relational database,
b) MongoDB[11] as the NoSQL database for the storage of the various datasets,
c) Elasticsearch[12] as the search optimisation and indexing engine,
d) MinIO[9] as the temporarily storage data lake,
e) Vault[10] for the storage of sensitive data, such as usernames, passwords, tokens, APIs Keys, etc.,
f) Ethereum[13] as the transaction ledger technology,
g) HDFS[14] as the distributed file system,
h) Gitlab[15] as the repository for algorithms.

5.1.1.3 Data Exploration Module

The Data Exploration Module (DEM) of the BEYOND Cloud based Platform enables users (such as data asset consumers and data brokers) to search for data assets of interest, browse and explore the results, in order to assess their potential and identify assets potential for acquisition. Overall, the Data Exploration Module addresses the data assets' discoverability of the BEYOND Cloud based platform; enabling users that are interested in acquiring or trading particular data assets, to perform search queries and view the relevant results (along with basic information about their content, their owner, their volume etc.,) as well as request access on the data of interest, based on the predefined access policies.

To fulfil its full set of functionalities, as shown in the following figure, the Data Exploration Module consists of two Data Storage & Retrieval components and one Security component. In respect to the first, these are a) the *Query Builder,* an intelligence search interface enabling data asset discoverability and offering data exploration capabilities, by allowing data consumers/brokers to search across all BEYOND data assets, browse the results and explore the data assets of-interest that are suitable for acquisition and b) the *DataSet Recommender,* which provides data

[11] https://www.mongodb.com/
[12] https://www.elastic.co/
[13] https://ethereum.org/en/
[14] https://hadoop.apache.org/
[15] https://about.gitlab.com/

recommendations to data asset consumers, based on the query parameters selected and predefined access policies. The security component included in the Data exploration Module namely, the *Access Policy Engine* which is responsible for assessing the applicable access policies of a data asset and identify which of the potential search results that the Query Builder has provided should be visible to the users. As such the Data Exploration Module, ensures that the potential data consumers can only view data assets, that they are eligible to acquire.

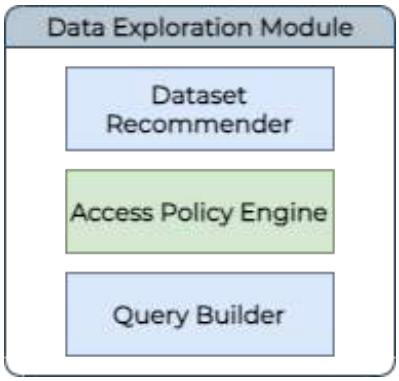

FIGURE 5-3: DATA EXPLORATION MODULE: COMPONENT OVERVIEW

General, through its security and retrieval/indexing mechanisms, the Data Exploration Module enables traceability of the data assets ingested in the BEYOND Cloud based platform, while ensuring their secure accessibility. Such functionalities are directly linked with the end-users' and business requirements as reported in D2.1 [2] and are interpreted here into a set of functional requirements that shall be met by the Data Exploration Module.

TABLE 5- 3 DATA EXPLORATION MODULE - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
DEM_1	The BEYOND solution shall enable users (i.e., data asset consumers and/or data brokers) to search for data assets available in the platform.	UC_D_4.9
DEM_2	The BEYOND solution shall provide a suitable searching mechanism supporting both keyword and/or text-based queries from the users.	UC_D_4.9
DEM_3	The BEYOND solution shall enable users to save their search queries and update them if needed at a later stage.	UC_D_4.9
DEM_4	The BEYOND solution shall enable users to apply filters on their search queries, based on the data asset's information (e.g., relevant domain, data type, data format, language, etc.).	UC_D_4.9
DEM_5	The BEYOND solution shall deliver an appropriate query definition mechanism providing a simple process for data asset searching (including AI models and analytics reports) available in the platform.	UC_D_4.9

DEM_6	The BEYOND solution shall provide a suitable searching mechanism enabling users to view the results (i.e., data assets) related to their search queries, along with the supplementary information; supporting an easier and faster identification.	UC_D_4.9
DEM_7	The BEYOND solution shall provide a suitable searching mechanism capable of providing recommendations of data assets or potential combinations of them, relevant to the users search queries, in case no relevant data asset is available in the platform.	UC_D_4.9
DEM_8	The BEYOND solution shall provide a suitable searching mechanism capable of recommending to data consumers, potential BEYOND stakeholders that could provide and/or create a data asset, relevant to their search query, in case no relevant asset is available in the platform.	UC_D_4.8
DEM_9	The BEYOND solution shall enable users to view a sample of the recommended data assets regardless of the assets' access policies.	UC_D_4.8
DEM_10	The BEYOND solution shall enable users to request access to a data asset they are not eligible to view/acquire from the relevant data provider.	UC_D_4.9
DEM_11	The BEYOND solution shall enable users to save the recommended data sets in their favourite lists, in order to be used again in the future for enhancing the recommendations results.	UC_D_4.8
DEM_12	The BEYOND solution shall enable users to view a list of all the recommendations for data assets and potential data providers, in case no asset relevant to their search query is available in the platform.	UC_D_4.8

Technology Stack:

The Data Exploration Module provide all the functionalities and services, extracted from the functional requirements as described in the previous section, through various state-of-the-art technologies that are presented below:

Back-end layer:

a) NestJS (NodeJS)[1] web framework.

Data Storage layer:

a) PostgreSQL,[8] for the storage of the queries configuration,
b) Elasticsearch[12], for performing the search queries in an effective and efficient manner.

5.1.1.4 Isolated Data Analytics Containers

The Isolated Data Analytics Containers (depicted in Figure 5-4) aim at providing services for extracting valuable insights and visualisations based on various analytics models and decision-making algorithms fed with in-house and third-party data ingested in the platform. In particular, data providers and consumers are provided with tools to design and execute advanced data analytics based on existing analytics models and algorithms (possibly pre-trained) to gain valuable information from their data and the corresponding analysis. These tools are provided through five core analytics components that are included in the Isolated Data Analytics Containers: (a) the Analytics Composer, (b) the PreTrained Analytics Sets, (c) the Analytics Executor, (d) the Visualisation Dashboard, and (e) the Results Exposer. Additionally, a Data Preprocessor component along with a Decryptor for decoding ingested data that are encoded, are included to serve as input blocks to the Analytics Composer component, while an Encryptor is used to encode potentially sensitive analytics results that are generated by the Results Exposer.

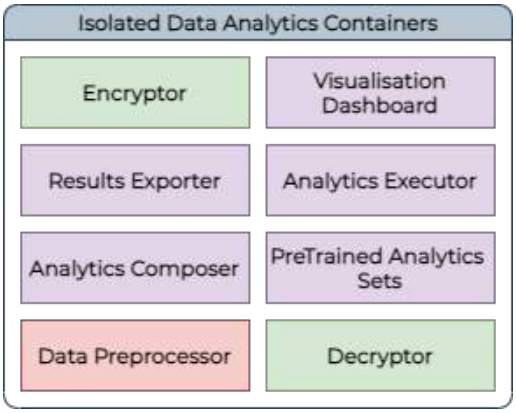

FIGURE 5-4: ISOLATED DATA ANALYTICS CONTAINERS: COMPONENT OVERVIEW

First, the Analytics Composer is responsible to provide a framework for designing analytics processes to the platform's data consumers. This framework requires the provisioning of data that is already processed in a particular format, in order to be fed through different analytics functions and algorithms that are configured appropriately by the data consumer, to comply with his/her needs. The analytics functions include, among others, machine learning algorithms such as regression, classification, clustering, forecasting, etc, supporting the decision-making capabilities of the BEYOND stakeholders. Moreover, the Isolated Data Analytics provides various pre-trained analytics models through the PreTrained Analytics Sets component that receives data that is already pre-processed, to facilitate the analytics design, as well as to reduce the configuration and execution time of analytics algorithms. Once the analytics workflows have been defined appropriately in Data Analytics or PreTrained Analytics Sets, the Analytics Executor is responsible to run the configured analytics workflows to receive the corresponding results. These results can be either visualised through the Visualisation Dashboard with which the data consumer is able to configure how the analytics results will be presented in the form of graphs, or through the Results Exposer with which the data consumer can export the analytics results to

the BEYOND End Users Tools to be used from the different BEYOND stakeholders' applications.

Functional Requirements

The Isolated Data Analytics Containers provide all the functionalities related to data analytics and visualisation, in a configurable framework that is defined by the various end-user and business requirements as reported in D2.1 "End-user & Business requirements analysis for big data-driven innovative energy services & ecosystems - a". Based on these requirements, the high-level functional requirements are extracted and listed below in Table 5- 4.

TABLE 5- 4 ISOLATED DATA ANALYTICS CONTAINERS - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
IDAC_01*	The BEYOND solution shall provide data preparation functions (e.g., sorting, appending, etc) prior their propagation to the various algorithms.	UC_D_4.1
IDAC_02*	The BEYOND solution shall allow users to design and configure the execution of analytics processes on their data.	UC_D_4.1
IDAC_03*	The BEYOND solution shall allow users to decode encrypted data such that appropriate data preparation functions to be configured and executed.	UC_D_4.1
IDAC_04*	The BEYOND solution shall allow users to add import blocks to propagate input datasets to data analytics processes.	UC_D_4.1
IDAC_05*	The BEYOND solution shall allow users to add manipulation blocks to apply preparation functions prior to data analytics processes.	UC_D_4.1
IDAC_06*	The BEYOND solution shall allow users to add analysis blocks to apply different data analytics and machine learning techniques over the imported data.	UC_D_4.1
IDAC_07*	The BEYOND solution shall allow users to add export blocks to allow the generation of the analytics results.	UC_D_4.1
IDAC_08*	The BEYOND solution shall provide an adequate list of AI algorithms (e.g., regression, classification, clustering, etc) to be utilised and applied on users' data.	UC_D_4.1
IDAC_09*	The BEYOND solution shall allow users to configure the execution parameters of the available AI algorithms.	UC_D_4.1

*** Denotes that a particular functional requirement is also elicited for the Private Polyglot Data Storage Layer.

IDAC_10*	The BEYOND solution shall provide pre-trained analytics models to be utilised for reduced analytics execution time.	UC_D_4.1
IDAC_11	The BEYOND solution shall execute all the data analytics processes in the BEYOND Cloud based Platform.	UC_D_4.1
IDAC_12*	The BEYOND solution shall enable users to select the format (e.g., csv, export for visualisation, etc) of the data analytics results.	UC_D_4.1
IDAC_13*	The BEYOND solution shall provide users a complete list of different visualisation graphs (e.g., line chart, bar charts, scatter plots, etc) for the visualisation of analytics' results.	UC_D_4.1
IDAC_14*	The BEYOND solution shall allow users to configure the visualisation of the analytics' results.	UC_D_4.1

Technology Stack

The Isolated Data Analytics Containers will provide the functionalities and services, extracted from the functional requirements as described in the previous section, through various state-of-the-art technologies listed below:

Back-end layer:

a) Nest (NodeJS)[1] web framework for delivering efficient, reliable, and scalable server-side applications,
b) Argo container native workflow engine,
c) Flask micro web framework for the exporting the analytics results,
d) Scikit-learn for analytics execution,
e) Spark, Pandas[3], and Dask frameworks for data manipulation,
f) RabbitMQ[5] message broker system for exchanging feedback messages, and keys to the Cloud Platform Operations Manager component.

Front-end layer:

a) VueJS[6] and TailwindCSS[7] for custom front-end design,
b) Kibana and/or Cube.js as the open-source data visualisation dashboards.

Data storage layer:

a) PostgreSQL[8] for storing and retrieving the configurations of data analysis processes, and data manipulation functions,
b) MinIO[9] for the temporarily storage of the transformed data,
c) HDFS[14] as the distributed file system,
d) Gitlab[15] as the repository for algorithms,
e) MongoDB[11] for storing the analytics results.

5.1.1.5 Cloud Platform Operations Manager

Overall, the Cloud Platform Operations Manager (CPOM) is shown in Figure 5-5 and consists of a bundle of security and system management components that collectively enable the efficient operation of the BEYOND Cloud based Platform. Such operations are mainly pertaining to different aspects and layers of security across the platform such as users' authentication and identity provider, monitoring of the available resources in the platform and most importantly orchestrating the operation and exchange of messages, data and controls between the BEYOND Cloud based Platform and the BEYOND Private Infrastructure (see 5.1.2).

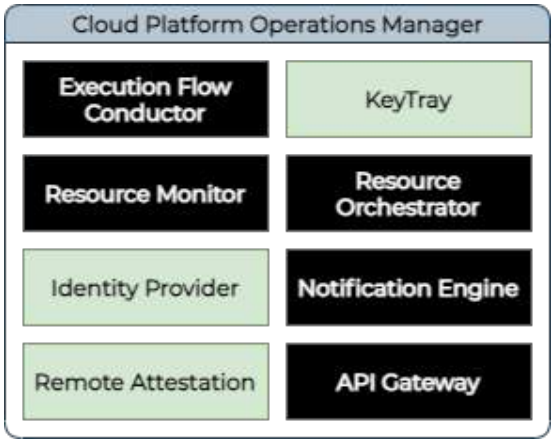

FIGURE 5-5: CLOUD PLATFORM OPERATIONS MANAGER: COMPONENT OVERVIEW

The system management components included in the Cloud Operations Manager are namely a) the Execution Flow Conductor, b) the Notifications Engine, c) the Resource Monitor, d) the Resource Orchestrator, and e) the API Gateway which are briefly described as follows.

The *Execution Flow Conductor* is responsible for the efficient orchestration of the various activities carried out in the Cloud based platform and the BEYOND Private Infrastructure along with their communication, by activating the relevant modules, layers and containers and feeding them with the information required for their successful operation. Moreover, the Execution Flow Conductor is responsible for the automatic execution of the different tasks associated with the data ingestion and data analytics activities, based on the schedule defined by the data asset owners or the data asset consumers, respectively, while also managing any errors that may occur during their execution. Additionally, the Execution Flow Conductor orchestrates the transfer of data assets that one organisation may own (or have purchased) from the BEYOND Cloud based Platform Storage to the BEYOND Private Infrastructure.

The Notifications Engine is responsible for providing appropriate real-time notifications to the BEYOND Cloud Based Platform's end-users, informing them about the various activities that take place in the platform, and which are of their concern. Such notifications may include updates on the progress of an ongoing data ingestion job (see 5.2.1 I) and/or a data analysis job (see 5.2.4 IV); as well as notify the responsible individuals of an organisation regarding the progress of data asset

brokerage activities at specific times. To minimise any delays in interpreting important notifications, (e.g., when a stakeholder request access to a data asset, or when a data asset contract is signed, rejected etc.), notifications are sent both via email to the respective stakeholders/responsible persons of organisations and through the BEYOND Cloud Based Platform.

The *Resource Monitor* enables monitoring of all the various data assets available in the BEYOND Cloud based Platform by providing insights of the added value of the various data assets and the usage of the different BEYOND Cloud based Platform components by combining and visualizing the appropriate statistics. Through the Resource Monitor, the BEYOND platform's administrators can gain a better understanding of the usage of the various BEYOND services and ensure their smooth operation by timely detecting problems and undertaking corrective actions or schedule maintenance activities. Moreover, Data Asset Providers can gain knowledge on the influence and popularity/usage of their data assets, as well as get a better view of all their data ingestion and data analysis jobs along with the associated resources.

The *Resource Orchestrator* is responsible for the smooth operation of the BEYOND Cloud based Platform, enabling the successful implementation of all services related to data ingestion and data analysis jobs, by allocating the appropriate computing and storage resources for the execution of each job, with regards to the volume of the data and the applicable services' requirements.

Lastly, the *API Gateway* represents the main entry point for any application/stakeholder (inclusive of all BEYOND End-User Tools) that needs to acquire data assets or analytics results which they are eligible to access from the BEYOND Cloud Based Platform. It is responsible for allowing the authorised stakeholders to retrieve data through the BEYOND Open APIs by using the same API method they had selected in the retrieval configuration, providing the unique identifier and the selected retrieval parameters. Whenever an authorised BEYOND stakeholder (or End-User tool) requests data, the API Gateway breaks it into multiple requests and directs them to the appropriate component (e.g., the data are retrieved from the Polyglot Data Storage layer, the applicable access policies are resolved in the Access Policy Engine, etc.) and generates a response.

The security components included in the Cloud Platform Operations Manager, are namely a) the KeyTray, b) the Identity Provider and c) the Remote Attestation, which are described in detail in section 5.1.1.7.

As previously mentioned, the Cloud Platform Operations Manager provides a variety of security and system management components aiming to ensure the smooth operation of the BEYOND Cloud based platform and its efficient interaction with the BEYOND Private Infrastructure. Under this context, Table 5- 5 provides a non-exhaustive list of functional requirements that shall be met by the Cloud Platform Operations Manager and which are directly linked with the end-users' and business requirements as reported in D2.1 [2].

TABLE 5- 5 CLOUD PLATFORM OPERATIONS MANAGER - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
CPOM_01*	The BEYOND solution shall provide an appropriate authentication engine capable of generating and securely managing the identity information of users belonging to an organisation and which are authorised to access the system.	UC_D_3.4
CPOM _02*	The BEYOND solution shall provide an appropriate authentication engine for the management of the registration process of organisations and for the validating the information submitted regarding their legal entity and responsible persons (e.g., the responsible manager for signing the data assets' contracts).	UC_D_3.4
CPOM _03*	The BEYOND solution shall provide an appropriate authentication and authorisation service to the end-users' tools for defining appropriate policies per tool, as well as generate and manage the relevant tokens.	UC_D_3.4
CPOM _04*	The BEYOND solution shall be able to manage the exchange of decryption keys, for the utilisation of encrypted data between the respective components of the platform ensuring end-to-end security.	UC_D_3.4
CPOM _05*	The BEYOND solution shall provide different mechanisms to verify the reliability and source of the data exchanged between the various components of the system.	UC_D_3.1
CPOM _06*	The BEYOND solution shall provide an appropriate notification mechanism able to identify the events taking place in the various components of the platform and their related recipients and generate the appropriate notifications to inform them.	No relevant UC
CPOM _07*	The BEYOND solution shall provide an appropriate notification mechanism able to generate and send appropriate notifications to the users regarding the execution progress of a data ingestion job (e.g., completion, failure, etc.).	No relevant UC
CPOM _08*	The BEYOND solution shall provide an appropriate notification mechanism able to generate and send appropriate notifications to the relevant users regarding the progress of a data asset brokerage activity (e.g., when it is due, when it is signed, rejected, etc.)	No relevant UC

*** Denotes that a particular functional requirement is also elicited for the Private Polyglot Data Storage Layer.

CPOM _09*	The BEYOND solution shall provide an appropriate notification mechanism able to generate and send the appropriate notifications to the relevant users both through the platform and/or via email.	No relevant UC
CPOM _10*	The BEYOND solution shall enable users to view and manage all the notifications related to their organisation.	No relevant UC
CPOM _11*	The BEYOND solution shall enable users to set their preferences on how to receive notifications (through the platform or via email) and based on their content.	No relevant UC
CPOM _12*	The BEYOND solution shall be able to dynamically orchestrate the available resources for the successful execution of each data check -in and/or analysis job.	UC_D_3.1
CPOM _13	The BEYOND solution shall be able to dynamically set and assign the appropriate computing and storage resources required for the successful operation of the platform and the efficient execution of its services.	UC_D_3.3
CPOM _14	The BEYOND solution shall be able to create and manage API keys for the different end-users' tools that have a retrieval scope and a specific expiry date.	UC_D_3.3
CPOM _15	The BEYOND solution shall be able to manage the different tasks included in receiving and processing simultaneous API calls, such as traffic management, version management, pagination and monitoring.	UC_D_3.3
CPOM _16	The BEYOND solution shall enable the platform administrators to monitor the platform's usage and services operations so as to plan maintenance activities, identify faults and timely carry out corrective actions.	UC_D_3.3
CPOM _17*	The BEYOND solution shall enable data asset providers to have an overview of their data assets' usage within the platform, as well as get an overview of all their data ingestion and analytics jobs.	UC_D_3.3
CPOM _18*	The BEYOND solution shall ensure that only anonymised data are collected towards extracting insights on the platform's usage, always adhering to the respective GDPR regulations.	UC_D_3.4

Technology Stack:

The Cloud Operations Manager will provide the intended functionalities and services as extracted from the aforementioned functional requirements, by leveraging various state-of-the-art technologies for the development of each subcomponent and which are described as follows:

Back-end layer:

a) Nest (NodeJS)[1] web framework for delivering efficient, reliable, and scalable server-side applications,
b) Keycloak[16], an open-source identity provider, for access management provision
c) Kubernetes[17] portable platform for managing containerised workloads and services in different Kubernetes clusters, as well as on Docker for containerisation of the different services and components of the BEYOND Cloud based platform,
d) Server-Sent Events (SSE) technology enabling a client to receive automatic updates from a server via HTTP connection.

Data Storage Layer:

a) PostgreSQL[8],
b) Elasticsearch[12] as the data storage mechanisms of the overall BEYOND platform.

5.1.1.6 Data Trading Module

The Data Trading Module (DTM), depicted in Figure 5-6, represents an one stop source of building-relevant (both energy and non-energy) data assets for stakeholders engaged in the energy value-chain; providing a secure and trusted environment for data asset sharing and trading. The Data Trading Module consist of various data trading components namely: a) the Contract Composer, b) the Trusted Distributed Ledger Technology (DLT) Engine, c) the BEYOND Data Marketplace, e) the Compensation Engine and f) the BEYOND Platform Wallet which are briefly described as follows.

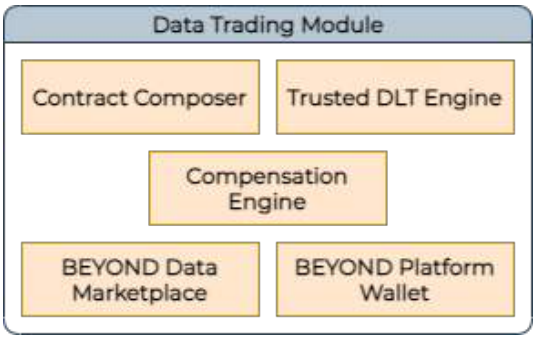

FIGURE 5-6: DATA TRADING MODULE: COMPONENT OVERVIEW

Through the BEYOND Data Marketplace, data consumers and/or data brokers are enabled to easily browse through and explore the available data assets in the Cloud based platform, preview their details and if satisfying their needs, proceed with their purchase by signing smart data asset contract with the respective data owner(s). Through communication with the Data Exploration Module (see 5.1.1.3), the BEYOND

[16] https://www.keycloak.org/
[17] https://kubernetes.io/

Data Marketplace, provides users with an appropriate interface enabling them to search for a data asset, see available data asset recommendations, browse through the results; as well as draft, negotiate and once agree on a suitable fee, sign a smart contract for the acquisition of a data asset of interest.

Such smart contracts are generated in the Contract Composer, supported by the Trusted DLT Engine. Overall, the Contract Composer is responsible for orchestrating the operations related to a smart data contract, including initiation of a data acquisition process, creation of a machine-processable smart contract format, users support during the drafting of the contract, secure storage of the contract in a distributed ledger through the Trusted DLT Engine and enforcement of the respective contract terms for data assets utilised by the BEYOND end-users tools and services.
For a signed contract to be activated, the data consumer/broker shall settle the relevant payment with the help of the Compensation Engine, which acts as an intermediary between the data asset consumer and the different data asset providers, responsible for processing and validating payments related to a smart data asset contract. To minimise the risk of any currency volatility, the Compensation Engine reimburses the data asset providers via cryptocurrency payments in their own wallets, utilizing Ethereum (ETH) as the chosen cryptocurrency.

The Compensation Engine also facilitates the calculation and reservation of the fee that shall be paid to the different data asset providers based on the respective signed smart contract and agreed fee for the acquisition of a data asset through the BEYOND marketplace. Lastly, through the BEYOND Platform Wallet, data asset consumers and / or data providers will be able to create a blockchain account (including their blockchain address) enabling them to come into contract with data asset providers. Through, the Platform Wallet users will be able to view their balance, their transaction history, and add/receive funds for the settlement (i.e., activation) of a data asset contract they have signed. It shall be noted that to enhance transparency and provide full traceability, all data asset contract-related information (e.g., negotiation steps, counteroffers, etc.) are stored in the BEYOND Transaction Ledger of the Polyglot Data Storage Layer (see 5.1.1.2).

The Data Trading Module through its various subcomponents will provide all the aforementioned functionalities related to data sharing and trading, in a configurable framework, that is defined by the various end-user and business requirements as reported in D2.1 [2].

Based on these requirements, the high-level functional requirements are extracted and listed below:

TABLE 5- 6 DATA TRADING MODULE - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
DTM_01	The BEYOND solution shall provide a trusted data sharing mechanism utilising a blockchain based data marketplace.	UC_D_4.5
DTM _02	The BEYOND solution shall provide an appropriate interface where data consumers/data brokers will be able to enter into multi-party contract with data owners for the acquisition of their data assets.	UC_D_4.4
DTM _03	The BEYOND solution shall provide an appropriate mechanism enabling users to draft smart asset contracts for the acquisition of data assets	UC_D_4.4
DTM _04	The BEYOND solution shall enable data owners to define their preferred license, IPR details and pricing scheme for their data assets.	UC_D_4.4
DTM _05	The BEYOND solution shall provide an appropriate mechanism enabling users to negotiate the terms and respective fee of data asset contract.	UC_D_4.4
DTM _06	The BEYOND solution shall provide an appropriate mechanism for evaluating any attempted data asset activity against the respective contract's terms to ensure that the terms are not violated.	UC_D_4.5
DTM _07	The BEYOND solution shall provide an appropriate mechanism for checking if a smart data contract is active and valid, prior to enabling access to the data consumer/broker.	UC_D_4.5
DTM _08	The BEYOND solution shall enable data consumer/brokers to create their own wallet to add funds, see their transaction history and pay the respective fees of a smart data asset contract.	UC_D_4.5
DTM _09	The BEYOND solution shall provide an appropriate remuneration mechanism, capable of managing the payments related to the activation of a smart contract.	UC_D_4.5
DTM_10	The BEYOND solution shall enable the parties involved in a smart asset contract to download the contract terms (e.g., pdf) for evaluation and archiving purposes.	UC_D_4.4

DTM_11	The BEYOND solution shall utilise a trusted cryptocurrency technology for the settlement of the smart data asset contracts	UC_D_4.5
DTM_12	The BEYOND solution shall enable data consumer/brokers to create their own wallet to add funds, see their transaction history and pay the respective fees of a smart data asset contract.	UC_D_4.5
DTM_13	The BEYOND solution shall enable data owners to create their own wallet, see their transaction history and receive the funds related to the activation of a smart data asset contract.	UC_D_4.5
DTM_14	The BEYOND solution shall enable also conventional payments (e.g., bank transfer) to the account of the data provider, or via the credit card of the data consumer.	UC_D_4.5

Technology Stack:
The Data Trading Module will provide the intended functionalities and services as extracted from the aforementioned functional requirements, by leveraging various state-of-the-art technologies for the development of each subcomponent and which are described as follows:

Back-end layer:

a) Nest (NodeJS)[1] web framework

Front-end layer:

a) VueJS[6],
b) TailwindCSS[7] for custom front-end design.

Data Storage layer:

a) Elasticsearch[12], for efficient searchability over the data and/or metadata

Blockchain layer:

a) Ethereum[13] distributed platform,
b) existing wallet providers such as MetaMask[18] for checking balance and transferring funds.

[18] https://metamask.io/

5.1.1.7 Security Components

The Security Components of the BEYOND Core Platform as mentioned before are: Encryptor, Decryptor, Anonymizer, KeyTray, Identity Provider, Remote Attestation, Access Policies Editor and Access Policy Engine.

In more detail:

The Encryptor component resides in the data ingestion services module and is responsible for providing the means for the data owner/producer to design the encryption policies (attribute-based encryption or ABE) to be applied on the uploaded data sets. Attributes are assigned to users and only the users that possess the appropriate attributes (e.g., country related attributes) can decrypt the corresponding data sets. This is performed by the generation of a unique key (generated and obtained from the KeyTray component) that is therefore assigned to the corresponding authorised users.

The KeyTray resides in the Cloud Platform Operations Manager and is responsible for the generation, storage, and lifecycle management of the encryption/decryption keys for the various datasets. According to the aforementioned ABE scheme, a predefined set of attributes produces a unique decryption key that will be used by the authorised users, e.g., data consumers that performed a raw/analytics data request to the core platform. On the opposite scenario, during the data ingestion phase, the data owners need to generate a unique encryption key for their data assets (if they choose so) before uploading them to the platform.

The Decryptor resides in the Isolated Data Analytics Containers module and is responsible for the decryption process of any encrypted data assets. This flow is triggered when a data consumer performs a search query through the Data Handling Module for a specific raw data asset or analytic method and in case these items are encrypted, the appropriate decryption keys are requested from the KeyTray so as to fulfill the decryption process.

The Anonymizer resides in the Data Ingestion Services module and is responsible to protect sensitive data by transforming any undesirable identifier that could reveal private information, defining certain anonymisation rules on the data under consideration. This step is very crucial as sensitive data can potentially point to a particular individual such as name, address, bank details etc. The user can choose the level of anonymisation on a particular column in a data set by categorizing the column as identifying, insensitive or something in between the two extremes. The data anonymisation methods used will preserve the k-anonymity of the individual, thus no set of attributes can be joined with other sources of information to reveal the individual's unique identity.

The Identity Provider resides in the Cloud Platform Operations Manager and is responsible for user identity access management operations (IAM). It checks the user identity and privileges on resources that the user might request such as access to the Beyond Core Platform or access to the Beyond Private Infrastructure. It also provides

the functionality to the BEYOND administrators to monitor the user activity, change a user's role etc.

The Access Policy Engine resides in the Data Exploration Module and is responsible to evaluate a policy relevant to finding/having access to a dataset. It checks the DLT if the specific data asset has been bought from the data consumer or not. If the data consumer has not bought an asset (no entry in the DLT) then the policy Engine checks the access policy provided with the asset against the data consumer attributes and presents this to him. If the attributes are not there, he cannot even see the asset. In case the data consumer has already bought the asset then the Access Policy Engine checks the DLT, if the attributes are correct, access is granted. If not, then the rejected attributes are returned in the backend to log this denied access request.

Finally, the Remote Attestation component resides in the Cloud Platform Operations Manager and is responsible for the integrity verification of the remote devices of the data owners and data consumers against the Beyond Cloud Platform. The data user's remote device plays the role of the prover whereas the platform has the role of the verifier which is a trusted party and checks whether the prover's software is to be trusted or not. So, any time a data owner or a data consumer attempts to connect to the Cloud Platform Operations Manager the software of his device is attested and the trustworthiness of the reported state of the software dictates if the connection attempt came from a trusted external (to the Beyond core ecosystem) device or not. As Remote Attestation might not be available in all devices (for example not all devices have TPM chips), this component will be an optional component that can be used by users.

Functional Requirements:
The components described above must fulfill specific functional requirements which are listed in the Table 5- 7 below:

TABLE 5- 7 BEYOND SECURITY COMPONENTS - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
SEC_001	The Beyond Security Components shall support the authentication of user identity.	UC_3.4
SEC_002	The Beyond Security Components shall allow the data uploading on authorised users/groups only.	UC_4.2
SEC_003	The Beyond Security Components shall support the encryption of uploaded data when necessary.	UC_4.2
SEC_004	The Beyond Security Components shall support the handling of the lifecycle of encryption keys.	UC_3.4
SEC_005	The Beyond Security Components shall provide a mechanism for securing encryption/ decryption keys.	UC_3.4
SEC_006	The Beyond Security Components shall provide end-to-end encryption of uploaded data based on hybrid	UC_4.2

	techniques (Attribute-Based Symmetric Searchable Encryption).	
SEC_007	The Beyond Security Components shall enable authorised users only, to retrieve information based on their data contract rules.	UC_3.4
SEC_008	The Beyond Security Components shall enable authorised users only, to acquire decrypted information based on their data contract rules.	UC_4.1
SEC_009	The Beyond Security Components shall enable authorised users only, to analyse/process information based on their data contract rules.	UC_4.1
SEC_010	The Beyond Security Components shall allow the data owners to pseudo anonymise or anonymise their data.	UC_4.1
SEC_011	The Beyond Security Components shall allow the data owners to define their own data anonymisation rules.	UC_4.1
SEC_012	The Beyond Security Components shall not allow the linking of a specific data set or data action with a specific actor.	UC_4.1
SEC_013	The Beyond Security Components shall support attribute-based access control models/policies (ABAC).	UC_3.4
	The Beyond Security Components shall provide attribute-based configuration/update on the configuration on the smart contracts signed between producer-platform and consumer-platform.	UC_3.4
SEC_014	The Beyond Security Components shall link the uploaded/analysed data to certain access control policies.	UC_3.4
SEC_015	The Beyond Security Components shall allow the data owners to define their own data access control policies at user/group level.	UC_3.4
SEC_016	The Beyond Security Components shall allow the data consumers to process the data according to the already specified policies stored in the Access Policy Engine.	UC_4.1
SEC_017	The Beyond Security Components shall not allow access/processing on data that have been identified as personal.	UC_4.1
SEC_018	The Beyond Security Components shall support the availability and usage of private trusted data containers to authorised users.	UC_4.2
SEC_019	The Beyond Security Components should support hardware-based remote attestation.	UC_3.4
SEC_020	The Beyond Security Components shall support software-based remote attestation when hardware-based attestation is not possible.	UC_3.4

Technology Stack:

The Beyond Security Components will provide the intended functionalities and services as extracted from the functional requirements, by leveraging various state-of-the-art technologies which are described as follows:

Back-end layer:

a) Keycloak, an open-source identity provider, for access management provision,
b) Casbin, an open-source authorisation library that supports access control models like ACL, RBAC, ABAC,
c) Vault for securing sensitive and secret parameters, distribution and key lifecycle management,
d) Google Tink, an open-source cryptography library that exposes simple and secure APIs,
e) Java for the encryptor and decryptor components,
f) Drools for the access policy editor component,
g) C++, Python and IBM Trusted Software Stack (TSS) for the remote attestation component.

5.1.2 BEYOND Private Infrastructure – Detailed Architecture

The BEYOND Private Infrastructure is intended to be installed at an organisation's premises to ensure secure data ingestion and analysis. In particular, the BEYOND Private Infrastructure consists of three core components: (a) the Private Data Ingestion Services which provides all functionalities related to data loading in the BEYOND Private Infrastructure, (b) the Private Polyglot Data Storage Layer which is responsible to store the ingested data and metadata in a secure space, and (c) the Data Analytics Container which provides all the data manipulation, analysis, and visualisation functionalities over the ingested data.

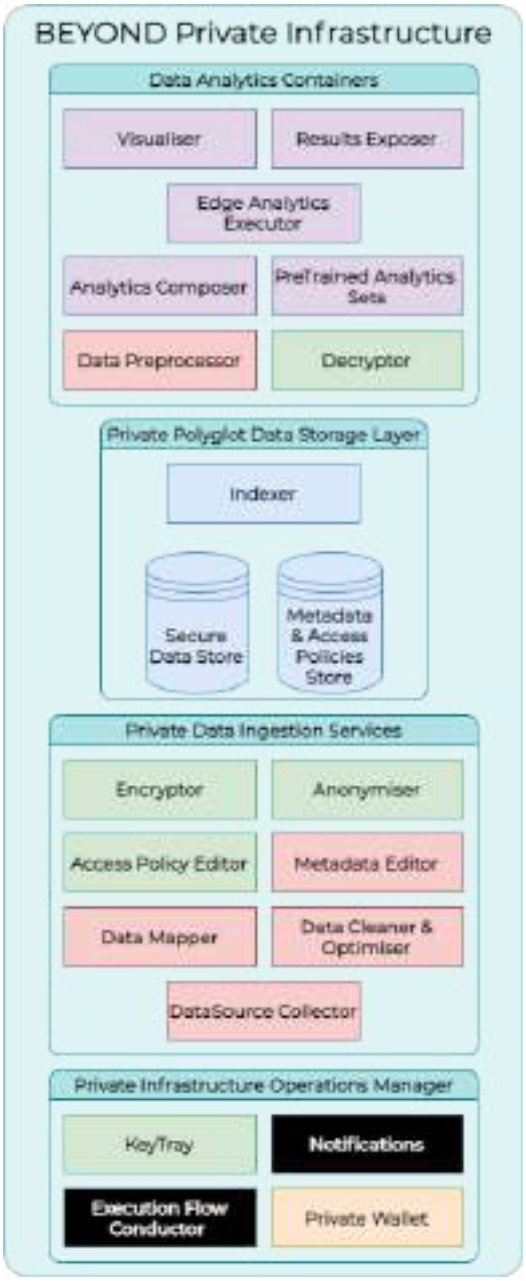

FIGURE 5-7: BEYOND PRIVATE INFRASTRUCTURE: CONCEPTUAL ARCHITECTURE

5.1.2.1 Private Data Ingestion Services

The BEYOND Private Infrastructure provides the Private Data Ingestion Services component which is deployed in the premises of an organisation. Although this major component for ingesting data is fundamentally similar to the Data Ingestion Services component that is included in the BEYOND Cloud based Platform, it is designed to be deployed in the BEYOND Private Infrastructure, and thus in the premises of an organisation, for ensuring increased security. As shown in Figure 5-8, it includes

several Data Management Components such as: (a) the DataSource Collector, (b) the Data Cleaner & Optimiser, (c) the Data Mapper, and (d) the Metadata Editor; as well as three main Security Components: (a) the Access Policies Editor, (b) the Encryptor, and (c) the Anonymizer.

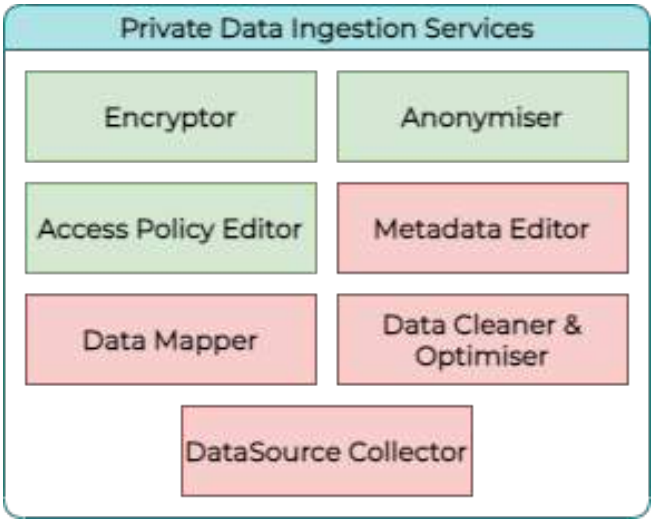

FIGURE 5-8: PRIVATE DATA INGESTION SERVICES: COMPONENT OVERVIEW

The DataSource Collector component is responsible to receive data collection configurations from the data providers in order to execute the data ingestion services using one of the available methods (i.e., Batch File Upload, API Request, or Subscription to Streaming Data). Data cleaning and optimisation services for ensuring that the ingested data are correct, consistent, and usable are provided by the Data Cleaner & Optimiser component which receives the cleaning configuration by the data providers and executes the cleaning processes to identify corrupted or missing data and replace them with appropriate values. The mapping of the ingested data to the common information model can be defined by data providers through the Data Mapper component which aims to establish a common understanding of the different concepts and fields within the BEYOND solution. The Metadata Editor of the Private Data Ingestion Services allows data providers to include metadata and any other additional description regarding their ingested data such that they are easily discoverable and potential relationships with other data could be established accordingly. It needs to be noted that the metadata added through the Metadata Editor component located in the BEYOND Private Infrastructure, are stored in the Metadata & Access Policies Store of the Private Polyglot Data Storage Layer which is deployed in the organisation premises for security.

The Access Policies Editor component is involved in the process of defining the policies for data access that allow or deny users from viewing or accessing particular data assets. Similarly, to the corresponding component deployed in the BEYOND Cloud based Platform, the different data access policies range from public, to restricted, and internal, denoting if the data is intended to be available to all, to other organisations who have signed data access contracts, or only to the data

owner/provider's organisation, respectively. Additionally, the Private Data Ingestion Services include another two security components, namely the Encryptor and the Anonymizer. The former is responsible to encode sensitive information that could be potentially carried by data assets. Moreover, the Encryptor generates a key through the KeyTray component deployed in the organisation's Private Infrastructure Operations Manager, in order to be used for decrypting the data when needed. The latter (Anonymizer) is responsible to remove any information from data that could potentially reveal private information identifying an individual.

Functional Requirements

Although the Data Ingestion Services of the BEYOND Cloud based Platform and the Private Ingestion Services of the BEYOND Private Infrastructure include components that provide similar functionalities, there is a clear separation between their deployment, configuration, and execution. Thus, the Private Data Ingestion Services component provides similar functionalities as the Data Ingestion Services that were also extracted mainly based on the end-user and business requirements as reported in D2.1 "End-user & Business requirements analysis for big data-driven innovative energy services & ecosystems - a".

TABLE 5-1: THE PRIVATE DATA INGESTION SERVICES - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
PDIS_01	The BEYOND solution shall allow data providers to upload their data to the BEYOND Private Infrastructure in different file formats (e.g., csv, json, etc).	UC_D_3.3
PDIS_02	The BEYOND solution shall support the execution of scheduled data uploading in the BEYOND Private Infrastructure.	UC_D_3.3

Additionally, the functional requirements of the Data Ingestion Services component (included in the BEYOND Cloud based Platform) that are followed by an asterisk, apply also for the case of the Private Data Ingestion Services, and thus they are not described in this section to avoid redundancy. However, for convenience of reference, the common functional requirements are mentioned by their ID here: DIS_03, DIS_05, DIS_08, DIS_09, DIS_10, DIS_11, DIS_12, DIS_13, DIS_14, DIS_15, DIS_16, DIS_17, DIS_18, DIS_19.

Technology Stack

The Private Data Ingestion Services provides all the functionalities and services, extracted from the functional requirements as described in the previous section, through various state-of-the-art technologies listed below:

Back-end layer:

a) Nest (NodeJS)[1] web framework for delivering efficient, reliable, and scalable server-side applications,
b) Flask[2] micro web framework,
c) Pandas[3] for handling, cleaning, and anonymizing data,
d) Kafka[4] distributed stream-processing software platform as the publication-subscription mechanism for streaming data collection,
e) RabbitMQ[5] message broker system for exchanging feedback messages, and keys to the Cloud Platform Operations Manager component.

Front-end layer:

a) VueJS[6],
b) TailwindCSS[7] for custom front-end design.

Data storage layer:

a) PostgreSQL[8] for storing the configurations of data ingestion, cleaning, anonymisation, encryption,
b) MinIO[9] for the temporarily storage of the ingested, cleaned, anonymised, encrypted, data and intermediate data management,
c) Vault[10] for securing sensitive and secret parameters.

5.1.2.2 Private Polyglot Data Storage Layer

The Private Polyglot Data Storage Layer (PPDSL), depicted in Figure 5-9, serves the data storage needs of the BEYOND Private Infrastructure and is deployed in the premises of an organisation for providing enhanced security. Its overall responsibility is the secure storage of all data assets (along with their associated metadata and access policies) that are ingested and/or generated in the BEYOND Private Infrastructure, while also making such data assets available to the BEYOND Cloud based Platform's modules and applications through direct communication with the Polyglot Data Storage Layer (see 5.1.1.2).

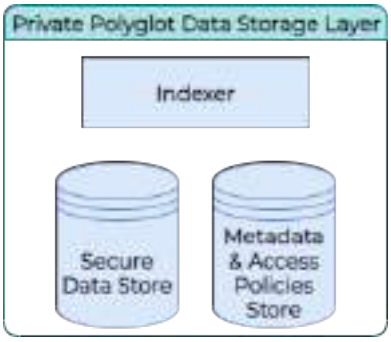

FIGURE 5-9: PRIVATE POLYGLOT DATA STORAGE LAYER: COMPONENT OVERVIEW

The Private Polyglot Data Storage Layer also enables the storage of all administrative information necessary for the efficient operation of the BEYOND Private Infrastructure

and regarding its internal structure, it is comprised of two storage containers namely, the *Secure Data Storage* and the *Metadata & Access Policies Store* and one data storage and retrieval component the *Indexer*. In more detail the *Secure Data Storage* is responsible for the persistent storage of all data assets generated or ingested in the BEYOND Private Infrastructure; while the associated metadata and access policies are configured in the Metadata Editor of the Private Data Ingestion Service (see 5.1.2.1), along with the links to the related data assets are stored *Metadata & Access Policies Store*. Finally, the *Indexer* is responsible for managing the indexing of the stored data assets (and their metadata) in the Private Polyglot Data Storage Layer so as to be easily searchable, while respecting the access policies set by their respective data providers.

It shall be noted that even though the Private Polyglot Data Storage Layer presents many similar components and analogous functionalities to the Polyglot Data Storage Layer of the BEYOND Cloud based platform, there is a clear separation between their design, deployment and implementation.

Functional Requirements

Although the Polyglot Data Storage Layer of the BEYOND Cloud based Platform and the Private Polyglot Data Storage Layer of the BEYOND Private Infrastructure include components that provide similar functionalities, there is a clear separation between their deployment, configuration, and execution. Thus, Private Polyglot Data Storage component provides similar functionalities as the Data Ingestion Services that were also extracted mainly based on the end-user and business requirements as reported in D2.1[2].

TABLE 5-2 PRIVATE POLYGLOT DATA STORAGE LAYER - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
PPDSL _01	The BEYOND solution shall provide a secure and trusted container for storing all the private data assets in the BEYOND Private Infrastructure.	UC_D_3.3
PPDSL _02	The BEYOND solution shall provide a secure and trusted container for storing all the private data assets metadata and access policies in the BEYOND Private Infrastructure.	UC_D_3.3

Additionally, the functional requirements of the Polyglot Data Storage Layer component (included in the BEYOND Cloud based Platform) that are followed by an asterisk, apply also for the case of the Private Polyglot Data Storage Layer, and thus they are not described in this section to avoid redundancy. However, for convenience of reference, the common functional requirements are mentioned by their ID here: PDSL_01, PDSL _02, PDSL _03, PDSL _04, PDSL _06, PDSL _07.

Technology Stack

The Private Polyglot Data Storage Layer provides all the functionalities and services, extracted from the functional requirements as described in the previous section, through various state-of-the-art storage and indexing technologies listed below:

a) PostgreSQL[8] as the BEYOND Private Infrastructure relational database,
b) MongoDB[11] as the NoSQL database for the storage of the various datasets,
c) Elasticsearch[12] as the search optimisation and indexing engine.

5.1.2.3 Data Analytics Container

The Data Analytics Container (shown in Figure 5-10) exploits the data ingested by the data owners/providers, to offer different solutions such as data processing, data manipulation, data analysis and decision making to different types of users withing an organisation. The main goal of the Data Analytics Container is to provide all the services needed for generating valuable information and visualisations based on various analytics models and decision-making algorithms fed with in-house and third-party data ingested in the platform. The Data Analytics Container consist of five Data Analytics components, namely: (a) the Analytics Composer, (b) the PreTrained Analytics Sets, (c) the Edge Analytics Executor, (d) the Visualiser, and (e) the Results Exposer, as well as a Data Preprocessor, and a Decryptor.

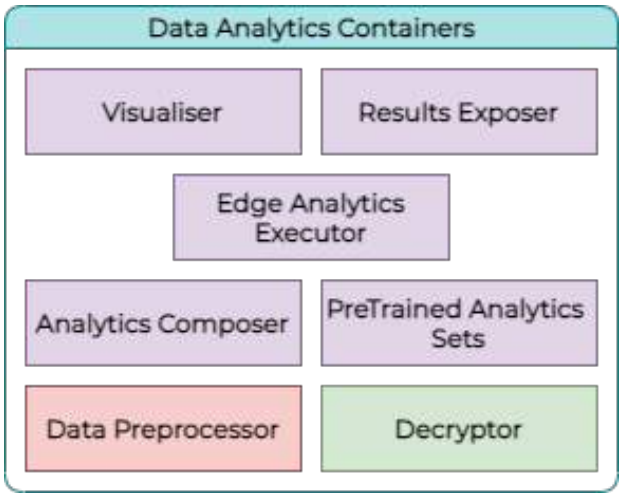

FIGURE 5-10: DATA ANALYTICS CONTAINER: COMPONENT OVERVIEW

The Analytics Composer provides a configurable framework to users for designing data analytics processes. In particular, the configuration of the data analytics processes involves the definition of the different blocks (i.e., data preparation, machine learning algorithms, etc) that will form a complete data analytics process, along with a particular schedule for the actual analytics execution. Users may also exploit several pre-trained analytics models through the PreTrained Analytics Sets component to facilitate the analytics design, as well as to reduce the configuration and execution time of their analytics processes. As soon as the data analytics processes have been

defined, their execution will be performed through the Edge Analytics Executor component, according to the configuration and schedule defined by the user. It needs to be highlighted that the execution is performed in the private infrastructure, hence increased security of the corresponding analysis is ensured. The results of the data analysis can be either visualised through the Visualiser component in the private infrastructure or through the Results Exposer component for further processing.

Functional Requirements

The functional requirements for the Data Analytics Container component are extracted based on the end user and business requirements as documented in D2.1 [2], and they are listed in Table 5-3.

TABLE 5-3 DATA ANALYTICS CONTAINER – FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
DAC_01	The BEYOND solution shall execute all the data analytics processes in the BEYOND Private Infrastructure.	UC_D_4.1
DAC_02	The BEYOND solution shall perform all data analytics functionalities in a secure framework.	UC_D_4.1

Additionally, the functional requirements of the Isolated Data Analytics Containers component (included in the BEYOND Cloud based Platform) that are followed by an asterisk, apply also for the case of the Data Analytics Containers, and thus they are not described in this section to avoid redundancy. However, for convenience of reference, the common functional requirements are mentioned by their ID here: IDAC_01, IDAC_02, IDAC_03, IDAC_04, IDAC_05, IDAC_06, IDAC_07, IDAC_08, IDAC_09, IDAC_11, IDAC_12, IDAC_13, IDAC_14.

Technology Stack

The Data Analytics Containers will provide the functionalities and services, extracted from the functional requirements as described in the previous section, through various state-of-the-art technologies listed below:

Back-end layer:

a) Nest (NodeJS)[1] web framework for delivering efficient, reliable, and scalable server-side applications,
b) Argo[19] container native workflow engine,
c) Flask[2] micro web framework for the exporting the analytics results,
d) Scikit-learn[20] for analytics execution,

[19] https://argoproj.github.io/
[20] https://scikit-learn.org/

e) Spark[21], Pandas[3], and Dask[22] frameworks for data manipulation,
f) RabbitMQ[5] message broker system for exchanging feedback messages, and keys to the Cloud Platform Operations Manager component.

Front-end layer:

a) VueJS[6] and TailwindCSS[7] for custom front-end design,
b) Kibana[23] and/or Cube.js[24] as the open-source data visualisation dashboards.

Data storage layer:

a) PostgreSQL[88] for storing and retrieving the configurations of data analysis processes, and data manipulation functions,
b) MinIO[9] for the temporarily storage of the transformed data,
c) HDFS[14] as the distributed file system,
d) Gitlab[15] as the repository for algorithms,
e) MongoDB[11] for storing the analytics results.

5.1.2.4 Private Infrastructure Operations Manager

The Private Infrastructure Operations Manager (PIOM) which is shown in Figure 5-11, is responsible for the efficient operation of the BEYOND Private Infrastructure installed at an organisations' premises; by providing security, system management and data trading components mainly pertaining to the efficient operation, as well as to the different security aspects and authentication layers of the Private Infrastructure's operation. In more detail the Private Infrastructure Operations Manager consist of:

a) the *Flow Conductor* responsible for orchestrating all the in the BEYOND Private Infrastructure;
b) the *KeyTray* responsible for the storage and management of the encryption/decryption keys for the various datasets residing in the BEYOND Private Infrastructure and are identified as encrypted data;
c) the *Notification Dashboard*, responsible for triggering real-time notifications to the BEYOND Private Infrastructure end-users, informing them about the various processes taking place and which are of their concern;
d) the *Private Wallet*, enabling data asset consumers and / or data providers to create a blockchain account locally, allowing them to enter into contract with data asset providers.

Through the Private Wallet, users will be able to view their balance, their transaction history, as well as add funds for settling a data asset contract they have signed

[21] https://spark.apache.org/
[22] https://dask.org/
[23] https://www.elastic.co/kibana/
[24] https://cube.dev/

(through communication with the Compensation Engine of the BEYOND Data Trading Module).

FIGURE 5-11: PRIVATE INFRASTRUCTURE OPERATIONS MANAGER: COMPONENT OVERVIEW

Functional Requirements

The functional requirements for the Private Infrastructure Operations Manager component are extracted based on the end user and business requirements as documented in D2.1 [2], and they are listed in Table 5-4.

TABLE 5-4 PRIVATE INFRASTRUCTURE OPERATIONS MANAGER - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant UCs
PIOM_01	The BEYOND solution shall provide a dedicated mechanism for the secure and encrypted storage of an organisation's wallet on the Private Infrastructure.	UC_D_3.3
PIOM_02	The BEYOND solution shall enable organisations to create and store their wallet in order to be engaged into data asset contracts.	UC_D_4.6
PIOM_03	The BEYOND solution shall enable users to import and their own existing Ethereum wallet if available.	UC_D_4.6
PIOM_04	The BEYOND solution shall enable organisations to view their wallet balances, transaction history and send/receive funds related to the compensation of data asset contract.	UC_D_4.5

Additionally, the functional requirements of the Cloud Platform Operations Manager component (included in the BEYOND Cloud based Platform) that are followed by an asterisk, apply also for the case of the Private Infrastructure Operations Manager, and thus they are not described in this section to avoid redundancy. However, for convenience of reference, the common functional requirements are mentioned by their ID here: CPOM_01, CPOM _02, CPOM _03, CPOM _04, CPOM _05, CPOM _06, CPOM _07, CPOM _08, CPOM _09, CPOM _10, CPOM _11, CPOM _12.

Technology Stack

The Private Infrastructure Operations Manager will provide the functionalities and services, extracted from the functional requirements as described in the previous section, through various state-of-the-art technologies listed below:

Back-end layer:

a) Nest (NodeJS)[1] web framework for delivering efficient, reliable, and scalable server-side applications,
b) Keycloak[16], an open-source identity provider, for access management provision,
c) Kubernetes[17] portable platform for managing containerised workloads and services in different Kubernetes clusters, as well as on Docker for containerisation of the different services and components of the BEYOND Cloud based platform,
a) Server-Sent Events (SSE) technology enabling a client to receive automatic updates from a server via HTTP connection,
b) MetaMask[18] for managing digital assets (e.g., view balance, transfer funds).

Data Storage Layer:

a) PostgreSQL[8],
b) Elasticsearch[12] as the data storage mechanisms of the overall BEYOND platform.

5.2 Processes View

The following processes that are presented in the following sub-sections are the main recognised "user trips" over the platform, showcasing the major workflows of how the BEYOND platform can be used. This of course does not constitute and exhaustive list of such workflows, but is presented in this chapter so that the reader can comprehend the main flow of operations that can be executed by a user.

5.2.1 I – Data Ingestion

Ia – Cloud Platform Data Ingestion

The BEYOND Cloud Platform allows data providers to configure the entire process of uploading their data in the cloud platform, along with the different pre-processing steps that are to be applied on the data prior the actual data ingestion execution. The entire process view of the data ingestion and pre-processing steps configuration and execution is depicted in the following figure.

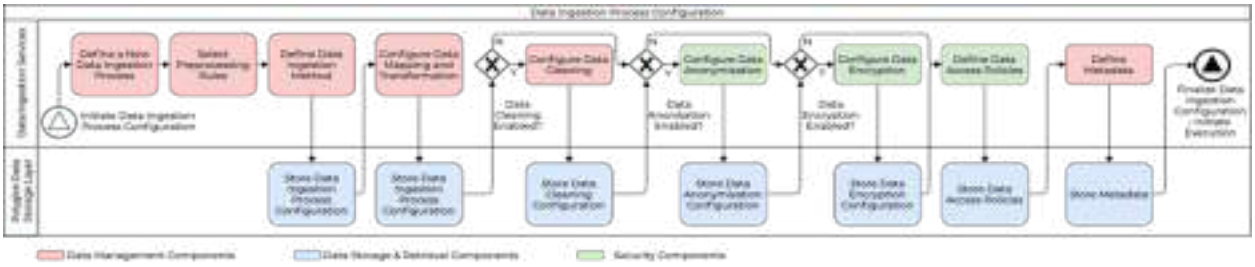

FIGURE 5-12: DATA INGESTION PROCESS

Data providers should create a new data ingestion process through the DataSource Collector, where the data ingestion method and the various data pre-processing steps are selected and defined. In particular, all the data pre-processing steps (i.e., data mapping, data cleaning, data anonymisation, and data encryption) that are to be applied on the ingested data, need to be selected prior their storage in the Polyglot Data Storage Layer. Upon the selection of the data pre-processing steps, the data provider shall define the data ingestion method ranging from batch file upload, API requests, to subscription to streaming data, along with a schedule defining when and how often to execute the data ingestion process, if applicable. As soon as the data ingestion method and its schedule has been defined, the data providers need to define and configure the various data pre-processing steps that have been selected before. First, data providers need to set the mapping from their source (ingested) data to the concepts and fields of the BEYOND CIM that is designed based on existing building-related and energy-related standards and located in the Polyglot Data Storage Layer from where is retrieved during the mapping configuration. The data mapping process is performed through the Data Mapper component which enables the definition of the transformation rules (i.e., measurement units, measurement scale, time zone) for numeric and datetime fields as well. Once the mapping and transformation rules are correctly defined and configured, the data provider may configure the remaining, yet optional, data pre-processing steps (i.e., data cleaning, data anonymisation, and data encryption).

The data cleaning rules can be defined through the Data Cleaner & Optimiser component which provides the framework to data providers in order to identify corrupted or missing data and configure appropriate methods for their replacement if needed. This is highly important to validate that the source data are correct, consistent and can be usable for further processing. Subsequently if the data anonymisation step is enabled, the data provider should provide the anonymisation rules to prevent data to be identified, through the Anonymizer component. Similarly, the data provider needs to define the encryption rules by selecting the parts of a dataset (or possibly the whole dataset) that are to be encrypted through the Encryptor component which receives an encryption key from the KeyTray component.

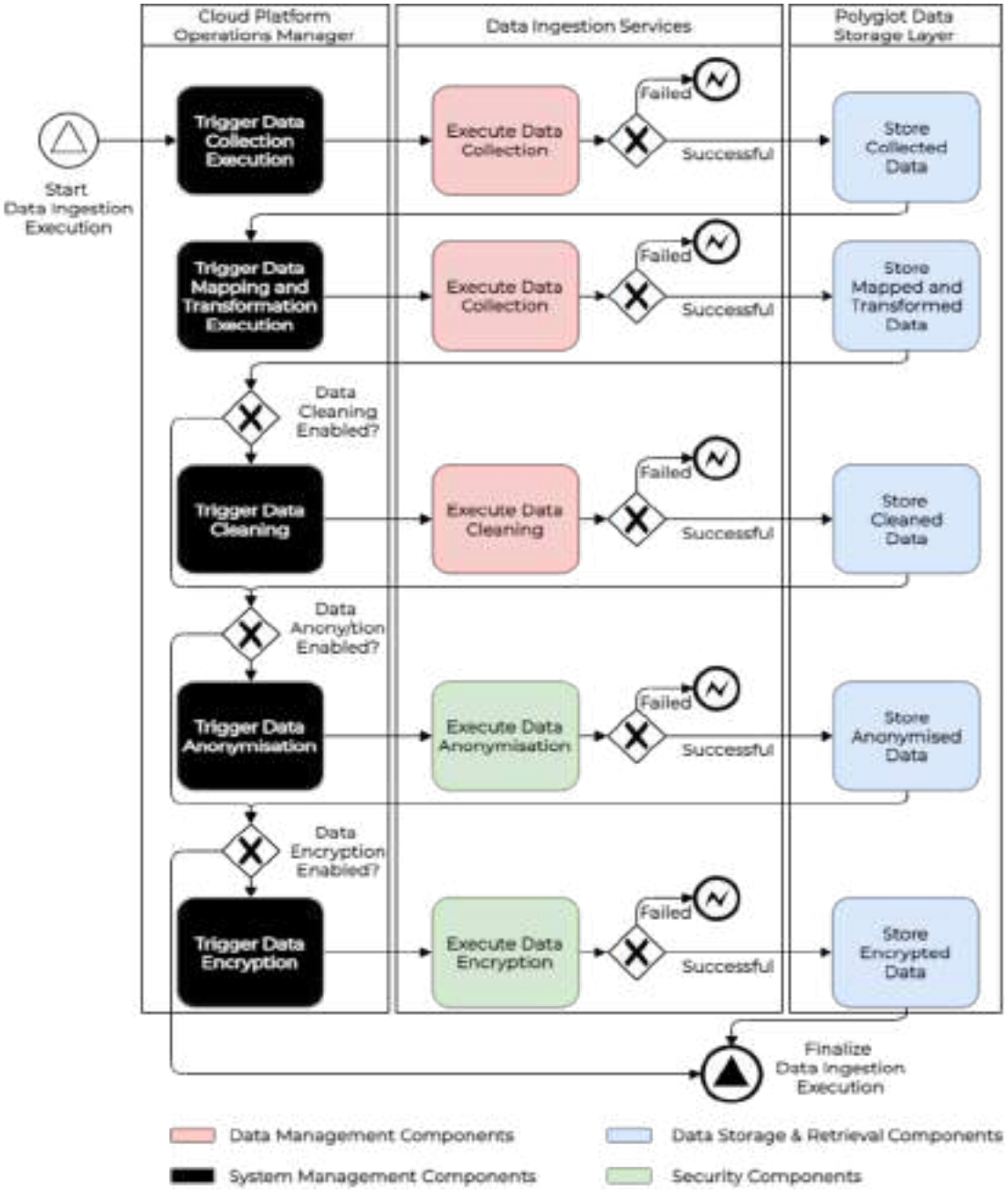

FIGURE 5-13: DATA INGESTION PROCESS EXECUTION

Finally, the data provider should set access policies to be applied on the data, through the Access Policies Editor, in order to deny or allow particular users to view or access the data. Along with the data access policies, the data provider should insert metadata and any information related to licensing through the Metadata Editor component, which proceeds in storing them in the Metadata & Access Policies Store of the Polyglot Data Storage Layer. Once all the necessary steps are defined successfully, the actual

execution of the different data ingestion processes, as depicted in Figure 5- 1, is scheduled to start according to the provided schedule.

The execution of the data ingestion process is directly triggered by the Execution Flow Conductor component of the Cloud Platform Operations Manager, based on the configured scheduled. It needs to be noted that since the data ingestion execution needs to start at a predefined date and time, the platform should have available resources to support the ingestion services. Thus, the Resource Monitor along with the Resource Orchestrator should collaborate in such a way to ensure that there exist allocated resources for the data ingestion services execution during the predefined period. As soon as the data ingestion execution has been performed successfully without any error during the entire process, the ingested data are stored temporarily, allowing the execution of the pre-processing steps (i.e., data mapping, data cleaning, data anonymisation, and data encryption).

In a similar way, the Execution Flow Conductor proceeds with the execution of the remaining pre-processing steps that have been scheduled and configured during the design time. In particular, the execution of each pre-processing step depends on successful execution of the predecessor step. Hence, a successful data mapping execution will result mapped and transformed data based on the BEYOND CIM, that could be potentially used in the data cleaning execution if selected as a pre-processing step by the data provider. Similarly, a successful data cleaning execution will produce a data asset cleaned from corrupted or missing data, that will enable the remaining pre-processing steps (i.e., data anonymisation, and data encryption) if enabled. Thereafter, the Execution Flow Conductor triggers the execution of the data anonymisation that will eventually provide a dataset that does not reveal any identifying information. Finally, the anonymised data are stored temporarily such that they are provided for the data encryption execution. During this step, the Encryptor requests an encryption key from the KeyTray which generates an encryption key (if not already generated) for the data that are to be encrypted.

It needs to be noted that in case of failure during the execution phase of the data pre-processing steps, appropriate messages notifying the data provider will be exchanged accordingly to identify the obstacles that prevent successful data ingestion and pre-processing steps execution. The data that were processed successfully by the different pre-processing steps, are transferred to the Polyglot Data Storage Layer.

Ib – Private Infrastructure Data Ingestion

The BEYOND Private Infrastructure allows data providers to configure the entire process of uploading their data in their private infrastructure, and to define all the necessary pre-processing steps that are to be applied on the data prior the actual data ingestion execution. The entire process view of the data ingestion and pre-processing steps configuration and execution follows similar flow as in the case of the Cloud Platform Data Ingestion described in the previous section. Data providers should

create a new data ingestion process through the DataSource Collector of the Private Data Ingestion Services, where the data ingestion method and the various data pre-processing steps are selected and defined. In particular, all the data pre-processing steps (i.e., data mapping, data cleaning, data anonymisation, and data encryption) that are to be applied on the ingested data, need to be selected prior their storage in the Private Polyglot Data Storage Layer. Upon the selection of the data pre-processing steps, the data provider shall proceed to select the files that are to be uploaded in the Private Infrastructure. As soon as the file that is to be uploaded has been selected and the execution schedule has been defined, the data providers need to define and configure the various data pre-processing steps that have been selected before.

First, data providers need to set the mapping from their source (ingested) data to the concepts and fields of the BEYOND CIM that is designed based on existing building-related and energy-related standards and located in the BEYOND CIM of the Polyglot Data Storage Layer from where is retrieved during the mapping configuration. The data mapping process is performed through the Data Mapper component, of the Private Data Ingestion Services, which enables the definition of the transformation rules (i.e., measurement units, measurement scale, time zone) for numeric and datetime fields as well. Once the mapping and transformation rules are correctly defined and configured, the data provider may configure the remaining, yet optional, data pre-processing steps (i.e., data cleaning, data anonymisation, and data encryption). The data cleaning rules can be defined through the Data Cleaner & Optimiser component which provides the framework to data providers in order to identify corrupted or missing data and configure appropriate methods for their replacement if needed. This is highly important to validate that the source data are correct, consistent and can be usable for further processing. Subsequently if the data anonymisation step is enabled, the data provider should provide the anonymisation rules to prevent data to be identified, through the Anonymizer component. Similarly, the data provider needs to define the encryption rules by selecting the parts of a dataset (or possibly the whole dataset) that are to be encrypted through the Encryptor component which receives an encryption key from the KeyTray component located in the Private Infrastructure Operations Manager. Finally, the data provider should set access policies to be applied on the data, through the Access Policies Editor, in order to deny or allow particular users to view or access the data. Along with the data access policies, the data provider should insert metadata and any information related to licensing through the Metadata Editor component, which proceeds in storing them in the Metadata & Access Policies Store of the Private Polyglot Data Storage Layer. Once all the necessary steps are defined successfully, the actual execution of the different processes within the Private Data Ingestion Services is scheduled to start according to the provided schedule.

The execution of the data ingestion process is directly triggered by the Flow Conductor component of the Private Infrastructure Operations Manager, based on the configured scheduled. As soon as the data ingestion execution has been performed successfully without any error during the entire process, the ingested data

are stored temporarily, allowing the execution of the pre-processing steps (i.e., data mapping, data cleaning, data anonymisation, and data encryption). In a similar way, the Flow Conductor proceeds with the execution of the remaining pre-processing steps that have been scheduled and configured during the design time. In particular, the execution of each pre-processing step depends on successful execution of the predecessor step. Hence, a successful data mapping execution will result mapped and transformed data based on the BEYOND CIM, that could be potentially used in the data cleaning execution if selected as a pre-processing step by the data provider. Similarly, a successful data cleaning execution will produce a data asset cleaned from corrupted or missing data, that will enable the remaining pre-processing steps (i.e., data anonymisation, and data encryption) if enabled. Thereafter, the Flow Conductor triggers the execution of the data anonymisation that will eventually provide a dataset that does not reveal any identifying information. Finally, the anonymised data are stored temporarily such that they are provided for the data encryption execution. During this step, the Encryptor requests an encryption key from the KeyTray which generates an encryption key (if not already generated) for the data that are to be encrypted. It needs to be noted that in case of failure during the execution phase of the data pre-processing steps, appropriate messages notifying the data provider will be exchanged accordingly to identify the obstacles that prevent successful data ingestion and pre-processing steps execution. The data that were processed successfully by the different pre-processing steps, are transferred to the Private Polyglot Data Storage Layer.

5.2.2 II – Data Exploration

Through the BEYOND Cloud based Platform, users such as data consumers and /or data brokers are enabled to address their specific data needs, by searching through the data assets ingested in the platform, or that that may belong to other BEYOND stakeholders (inclusive of End-User Tools), towards identifying potential ones for acquisition.

As shown in the next figure, depicting the overall data exploration process, data consumers and/or data brokers can address their specific data needs, by searching and exploring data assets (that they are authorised to view/acquire) and are available in the Cloud based Platform. When no relevant data assets are available, users are provided with recommendations of relevant data assets that can cover their needs, or data providers that can deliver the required data assets.

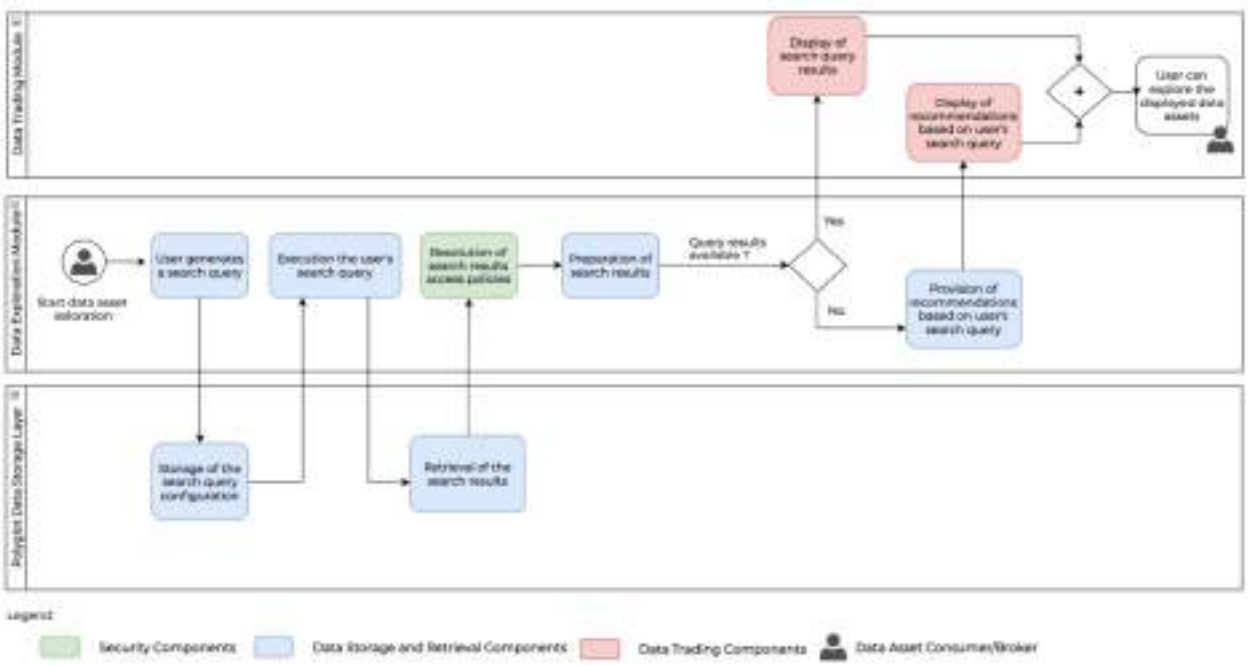

FIGURE 5-14: DATA EXPLORATION PROCESS

The process starts with the user creating a new search query (or loading one they had configured in the past) through the Query Builder. The configuration of the new search query is stored in the Polyglot Data Storage Layer, while the search query is executed both over the metadata and/or the data of the data assets stored in the Polyglot Data Storage Layer. Prior the Query Builder preparing the results for visualisation, the Access Policy Engine resolves the access policies of each of the results. When the search results do not breach any access policies, these are displayed to the data consumers/brokers supplemented by recommendations (provided via the DataSet Recommender) for data assets or potential data asset providers that can offer such data upon request. Once the search results are visible, users can filter and sort the results based on the different parameters offered, as well as explore the data samples of the resulting data assets.

If a data asset required, falls within the private or recommended data assets/providers, users can initiate the data trading process (as described in section 5.2.3), for acquiring the data asset they need through signing a smart asset contract with the respective data provider.

5.2.3 III – Data Trading

As described in the previous section (5.2.2 II), in the event a data asset required from a data consumer and/or data broker falls within the private data assets available in the cloud platform, or in the recommended data asset/data providers, they can initiate

the process for acquiring them by requesting a quote and signing a smart asset contract with the respective data provider.

As shown in the following figure, depicting the overall Data Trading process, through the BEYOND Data Marketplace data consumer and/or brokers can explore the resulting data assets of their search queries; if a data asset fulfils their requirements, they can request a quote for its acquisition from the respective data provider. Upon receival of the request notification, the data provider via the BEYOND Data Marketplace can view the quotes details and either reject the request (notifying the data consumer/broker) or accept it and proceed with drafting a smart data asset contract supported by the Contracts Composer and the Trusted DLT Engine and stored in the BEYOND Transaction Ledger of the Polyglot Data Storage Layer.

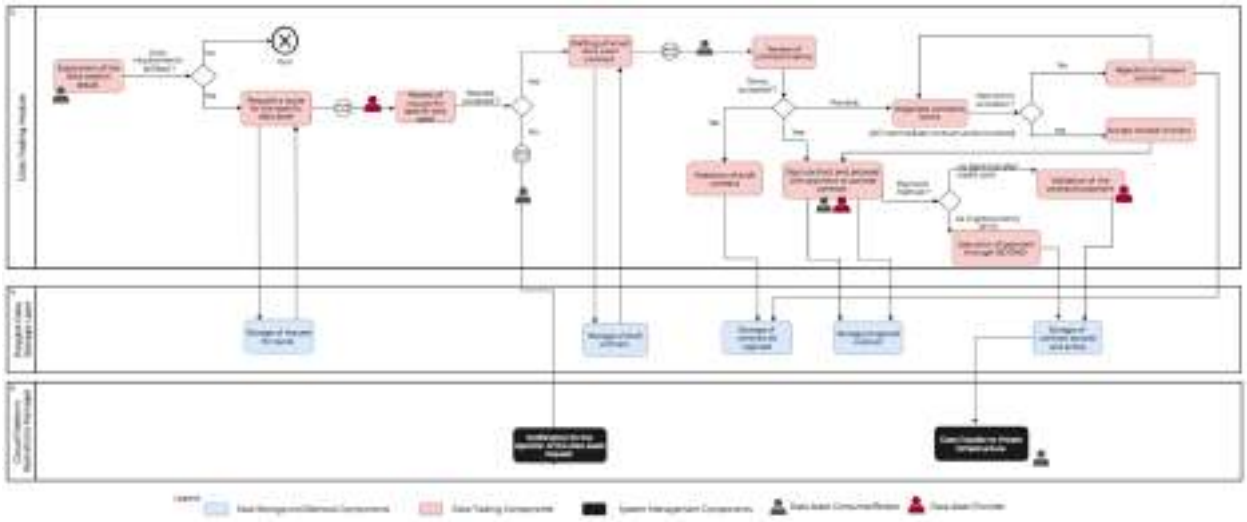

FIGURE 5-15: DATA TRADING PROCESS

Upon drafting the respective data asset contract, this is stored in the BEYOND Transaction Ledger of the Polyglot Data Storage Layer; while the data asset consumer is notified in order to review its details by accessing the BEYOND Data Marketplace. In the event the terms of the contract are acceptable, the data asset consumer can proceed with signing the contract and proceed with the payment; or he can reject the contract if the terms are not acceptable. In the event the data asset consumer partially accepts the terms of the contract, he/she can enter into negotiations with the respective data asset provider by providing a counteroffer (a revised draft contract) stored as a new version of the contract. Once the negotiations finish both the data asset provider and the data asset consumer shall accept or reject the latest version of the contract and if accepted from both, the latter is notified to proceed with the respective payment.

The payment can be made both through the BEYOND Cloud based platform's payment mechanism in cryptocurrency (with the help of the BEYOND Platform Wallet and the Private Wallet installed in each stakeholder's private infrastructure); or

via standard means of payment, such as a bank transfer or credit card payment to the data provider's account. When the payment is made via standard means, the data asset provider shall verify the receipt of the payment in the Compensation Engine, in order for the smart asset contract to become active. Finally, once a contract is marked as active, the respective data asset is automatically transmitted to the Private Infrastructure of the data asset consumer with the support of the Cloud Platform Operations Manager orchestrating the overall transfer of the respective data asset.

The Data Trading process described above entails the acquisition of a data asset belonging to one only data provider. However, there might be the case that a data asset has been generated as the outcome of a data analysis process (described in 5.2.4) combining various trained analytics models and datasets, that belong to multiple data asset providers. In this context, the Data Trading process for the acquisition of multiparty data assets is shown in and described as follows.

Through the BEYOND Data Marketplace data consumer and/or brokers can explore the resulting data assets of their search queries; if a data asset fulfils their requirements, they can initiate the process for acquiring it by requesting a quote from the respective data provider(s). The request is stored in the Polyglot Data Storage Layer and if the requested data asset belongs to multiple data asset providers, individual notifications for the request are to each one of them.

Via the BEYOND Data Marketplace, the involved data providers can review the request's details, where three types of actions are possible:

- **a)** all the data providers reject the request, and the data consumer/broker is notified.
- **b)** some of the data providers accept the request; in this case the data consumer's preferences defined during the request are checked, in the event they request the whole data asset, in such case the request is automatically rejected in the Data Marketplace. However, if only a part of the data asset is requested and no rejection is expressed by the respective data providers, without violating any access policies (e.g., product-data), then the request is automatically accepted in the Data Marketplace and drafting of a multiparty data asset contract between the data providers and the data consumer/broker can take place (as described below).
- **c)** all the data providers accept the request. In this case the data providers shall come into contract with BEYOND platform for sharing a data asset they are involved in by signing a smart data asset contract. A draft smart contract is generated by the platform and all involved data asset providers need to review its terms and decide whether to accept and sign it, reject it, or partially agree with its terms and negotiate with the involved parties. To provide full traceability of the brokerage process, each intermediate contract (counteroffer) is stored in the Trusted DLT Engine and once all data providers have responded the Contract Composer is responsible for checking if any data provider has

rejected the draft contract. If some have rejected it, the data consumer's preferences defined in the request are checked; in case they have requested the whole data asset or none, then the smart contract that has been signed by some data asset providers is automatically considered as cancelled. However, if any part of the data asset is requested and no access policies are breached (in case of product data), then drafting a multi-party data asset contract preparation between the data providers and the data consumers can take place. It should be noted that the drafting of a multi-party data asset contract can only take place if all the data asset providers have signed the smart contract with the BEYOND platform.

Drafting of the multi-party data asset contract is carried out by the Contract Composer combining the smart contract that is to be signed between the data asset consumer and the BEYOND platform that acting as an intermediary for the data asset providers to simplify the process and increase trust. Once the data asset consumer reviews the terms of the multiparty data asset contract, he/she may accept or reject them. No option for negotiation of the multiparty contract terms is envisioned for the data asset consumer in this stage of the design process, due to the many data asset providers involved, making the overall process very complicated.

Lastly, once a multiparty data asset contract is signed by the respective data consumers, they need to pay the corresponding fee in the Compensation Engine. In this case the payment can be only executed through the BEYOND Cloud Based platform, from the Private Wallet of the data asset consumer to the BEYOND Platform Wallet in the supported cryptocurrency (ETH). The data asset providers are then appropriately compensated based on the terms of the signed smart contract and the funds are transferred to their Private Wallet with the help of the Compensation Engine.

As soon as the payment and reimbursement process is validated, the multiparty data asset contract is stored as paid and active. The respective data assets are then automatically transferred to the data asset consumer's private infrastructure with the help of the Execution Flow Conductor coordinating the overall transfer process.

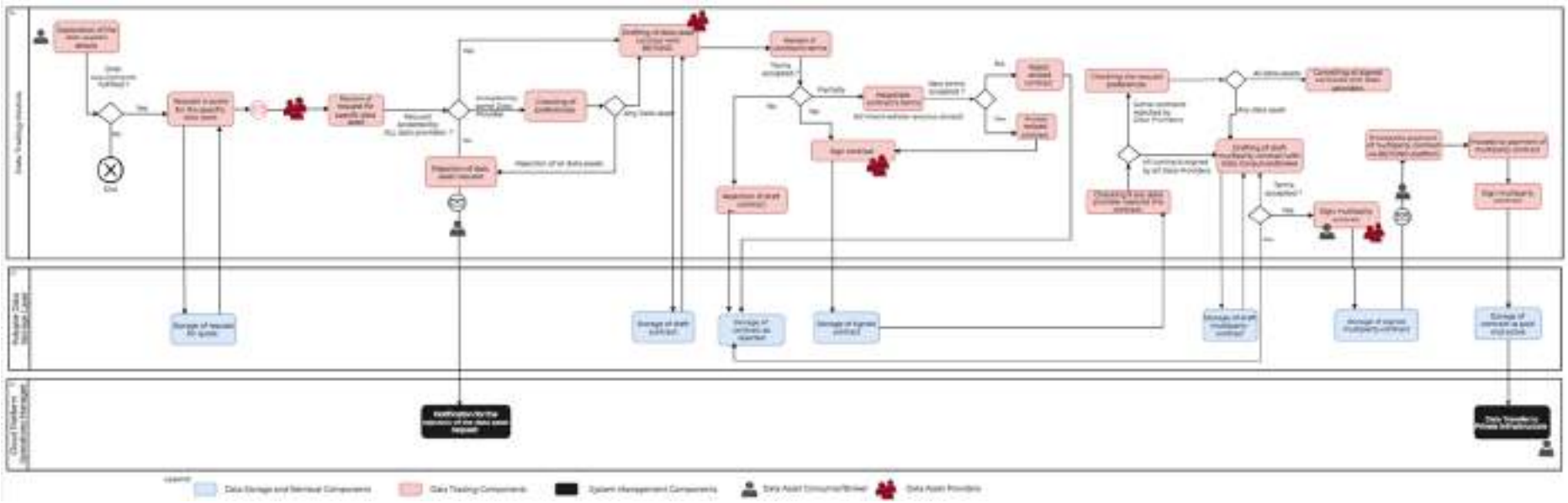

FIGURE 5-16: DATA TRADING PROCESS – MULTIPARTY DATA ASSET ACQUISITION

5.2.4 IV - Data Analysis

Data consumers are provided with data analytics capabilities that could be run either in the BEYOND Cloud based Platform or in the BEYOND Private Infrastructure. The analysis could be done over their own data loaded during the data ingestion phase by them, or over data that were loaded in the platform by other data providers, as described in Section 1. In this section the entire process of designing and executing data analytics is described towards the provisioning of the BEYOND platform for the different business problems regarding optimising energy policy planning, infrastructure sizing, de-risking renovation investments and energy performance management for buildings.

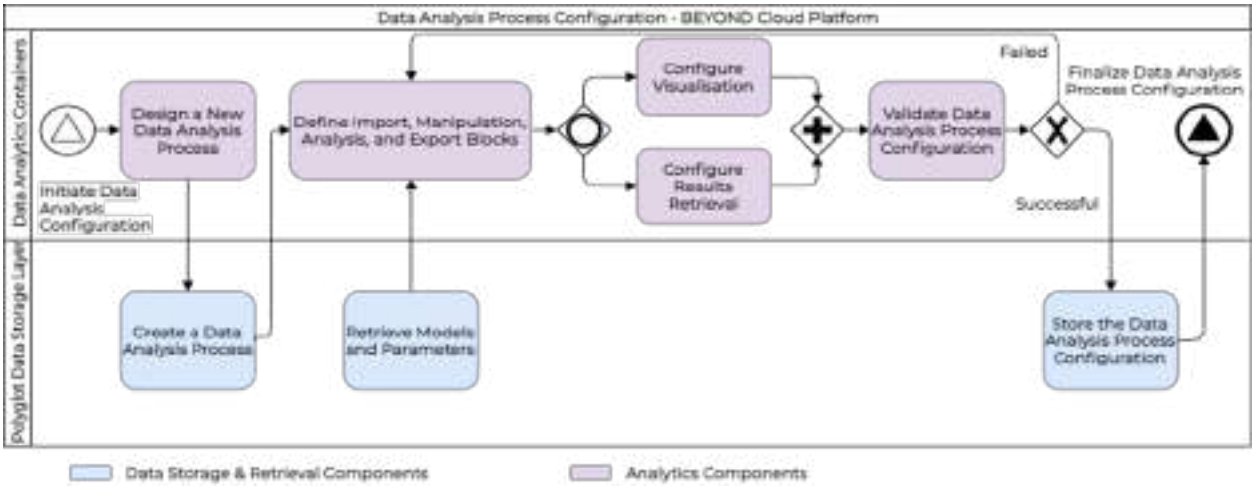

FIGURE 5-17: DATA ANALYSIS PROCESS - CONFIGURATION

A new data analytics process can be designed through the Isolated Data Analytics Containers which mainly receives execution commands from the Cloud Platform Operation Manager, and processed data from the Polyglot Data Storage Layer. A data analytics process consists of four different fundamental blocks that could be combined and placed accordingly to describe a complete data analytics process for a particular business problem. These blocks can be considered as categories of functions that contribute to the decision-making process of the data analysis. A complete data analytics process includes the following blocks:

1. Import Block, defined by data consumers to specify the datasets that are to be used as input for the various preparation and analytics functions. Data consumers are able to select only the datasets for which they have access.
2. Manipulation Block, defined by data consumers to specify the preparation and manipulation functions (e.g., sorting, appending, etc) that are to be applied on the imported data.
3. Analysis Block, defined by data consumers to specify the data analytics models including statistical models, machine learning algorithms, as well as pre-trained models (provided through the PreTrained Analytics Sets component).

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement n° 957020.

4. Export Block, defined by data consumers to specify how the results (that are expected to be obtained from the Algorithm Block) can be used for further analysis through the Results Exposer, or for visualisations configured by the data consumer through the Visualisation Dashboard.

These blocks can be used in such a way to form a complete data analytics process for a particular business problem. A data consumer is able to use multiple blocks where each block can use and propagate the output of its preceding block, and thus the placement of each block needs to be set carefully in order to prevent any execution failure.

IIIa – Cloud Platform Data Analysis

The execution of a data analysis process in the Isolated Data Analytics Containers of the BEYOND Cloud based Platform starts according to the defined schedule provided by the data consumer. Eventually, the Analytics Executor coordinates and triggers the execution of the data analytics blocks along with any supporting service needed for a data analysis process. In particular, the Analytics Executor verifies that the data that were configured as input in the import blocks are available in the Cloud based Platform. Following that, the Analytics Executor ensures that the Cloud based Platform has the available computational resources to run the defined manipulation blocks, through interaction with the Resource Monitor. In case that there are no available resources, it requests for dynamic allocation of resources from the Resource Orchestrator, in order to complete the different data manipulation tasks. As soon as the data manipulation blocks have been executed successfully, the processed data are stored temporarily for further analysis, otherwise the data analysis process fails to propagate the processed data to the following blocks. Similarly, for the analysis blocks, the Analytics Executor needs to ensure there are available computational resources to complete the data analytics functions as configured in the analysis blocks. The successful execution of the analysis blocks enables the execution of the following blocks by propagating their corresponding analysis results, otherwise the data analysis process fails to propagate the analytics results to the following blocks. Finally, for the export blocks, the Analytics Executor triggers the execution of either the Visualisation Dashboard or the Results Exposer, depending on the data consumer's choice during the data analysis process definition. If the data consumer selected to visualise the results directly in the Cloud based Platform, the Visualisation Dashboard will be executed and the raw data including the data analysis results will be stored in the Polyglot Data Storage Layer, otherwise if the selection was to export the results as a file, the Results Exposer will be executed to prepare the file and provide the corresponding link to download it accordingly.

The execution workflow of the data analysis process that runs in the BEYOND Cloud based Platform is depicted in the following figure. It needs to be noted that for the case where the data analysis process has been configured to be executed in the BEYOND Private Infrastructure, a similar workflow is followed, although the execution is performed in the Data Analytics Containers of the on-premise environment, and the

results from the different data analysis function blocks are stored in the Private Polyglot Data Storage Layer.

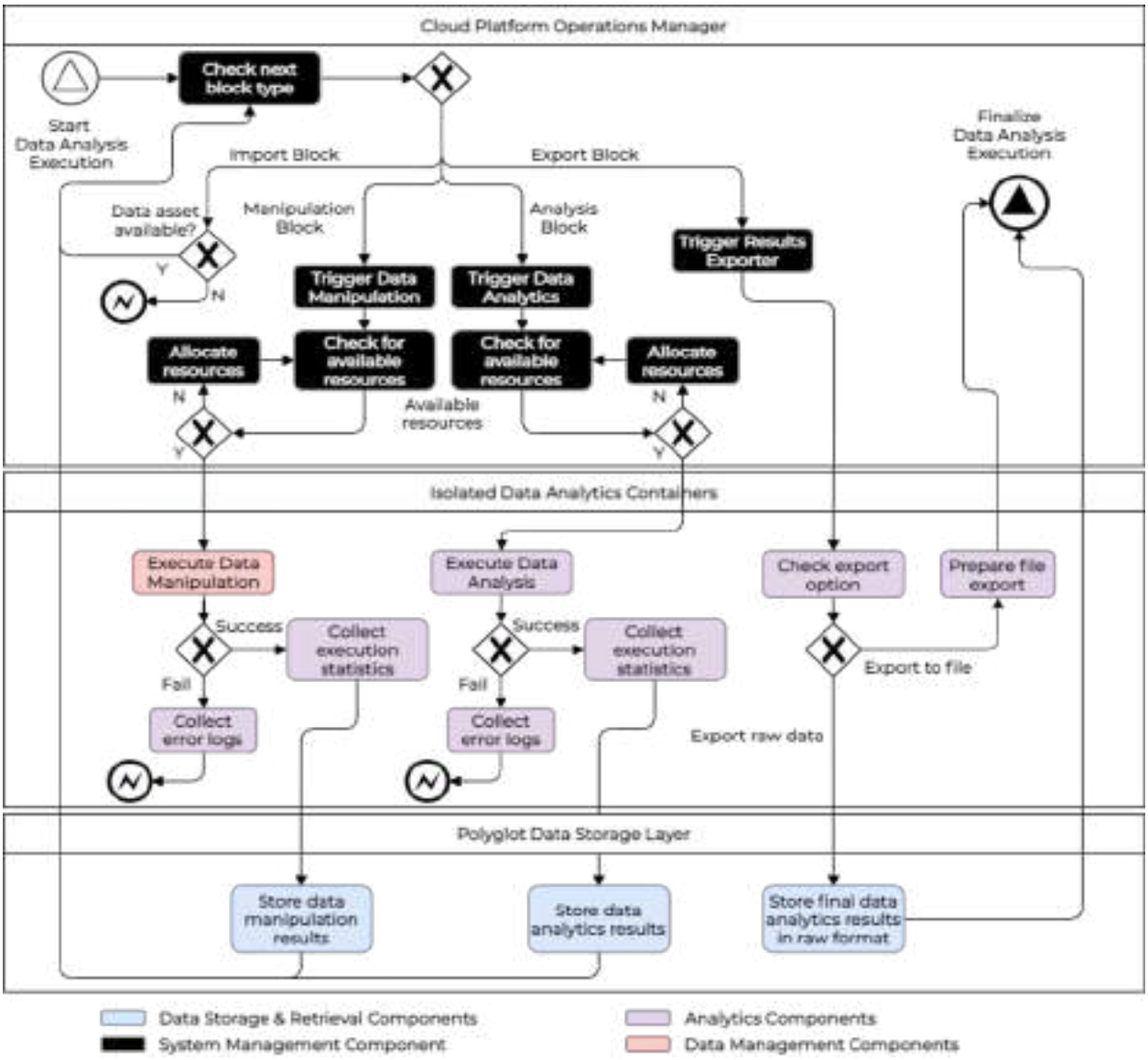

FIGURE 5-18: DATA ANALYSIS PROCESS - EXECUTION

IIIb – Private Infrastructure Data Analysis

If a data consumer selected to perform a data analysis process in the BEYOND Private Infrastructure, the entire process is set to be executed in the Data Analytics Containers. As already mentioned, the data analysis in the Private Infrastructure follows a similar path for its configuration and execution, as in the case of the Isolated Data Analytics Containers of the BEYOND Cloud based Platform, described in the previous section. According to the defined schedule, the Edge Analytics Executor coordinates and triggers the execution of the data analytics blocks along with any

supporting service needed for a data analysis process. In particular, the Edge Analytics Executor verifies that the data that were configured as input in the import blocks are available in the Private Infrastructure, otherwise it triggers the appropriate services to transfer them accordingly through the Private Data Ingestion Services. Following that, the Edge Analytics Executor triggers the different data manipulation tasks, as defined within the data manipulation blocks. As soon as the data manipulation blocks have been executed successfully, the processed data are stored temporarily for further analysis, otherwise the data analysis process fails to propagate the processed data to the following blocks. Similarly, for the analysis blocks, the Edge Analytics Executor triggers the data analytics functions as configured in the analysis blocks through the Analytics Composer and the PreTrained Analytics Sets (if the configured analysis makes use of pre-trained models). The successful execution of the analysis blocks enables the execution of the following blocks by propagating their corresponding analysis results, otherwise the data analysis process fails to propagate the analytics results to the following blocks. Finally, the Edge Analytics Executor triggers the execution of either the Visualiser or the Results Exposer, based on the data consumer's selection during the data analysis process definition. If the data consumer selected to visualise the results directly in the Private Infrastructure, the Visualiser will be executed to present the results graphically, otherwise if the selection is to export the results as a file, the Results Exposer will be executed to prepare the file and provide the corresponding link to download it accordingly.

6 BEYOND Core Platform Functional & Non-Functional Requirements

6.1 List of BEYOND Functional requirements

Having described the different processes undertaken within the BEYOND Core Platform and the respective workflows, this section summarises the list of functional requirements for the BEYOND Platform, as extracted from its individual components.

As shown in , such requirements are mapped (where applicable) to the respective business requirements of the BEYOND solution, as defined in D2.1 [2]. It shall be noted that the list of functional requirements for the various security components of the BEYOND platform are provided in Table 5- 7, thus excluded from the table below.

TABLE 6-1 BEYOND PLATFORM - FUNCTIONAL REQUIREMENTS

ID	Description	Relevant Business Req_id [2]
DIS_01	The BEYOND solution shall allow data providers to upload their data to the BEYOND Cloud based Platform in different file formats (e.g., csv, json, etc).	Req_016
DIS_02	The BEYOND solution shall allow data providers to load streaming data to the BEYOND Cloud based Platform.	Req_016
DIS_03*	The BEYOND solution shall support the temporal configuration of the data loading services (e.g., schedule data loading).	Req_017
DIS_04	The BEYOND solution shall support the execution of scheduled data uploading in the BEYOND Cloud based Platform.	Req_018
DIS_05*	The BEYOND solution shall support the execution of scheduled and periodic data loading.	Req_017 Req_018
DIS_06	The BEYOND solution shall support the execution of scheduled collection of streaming data.	Req_016
DIS_07	The BEYOND solution shall support the execution of scheduled collection of data from API.	Req_019
DIS_08	The BEYOND solution shall validate that the ingested data are of appropriate format, range, and content.	Req_001
DIS_09*	The BEYOND solution shall allow data providers to set rules for handling incomplete or corrupted entries, and outliers in their ingested data.	Req_001

* Denotes that a particular functional requirement is also extracted in the Private Data Ingestion Services component.

DIS_10*	The BEYOND solution shall allow data providers to set rules for filling empty entries (possibly detected) in their ingested data.	Req_003
DIS_11*	The BEYOND solution shall support the fusion of different data formats into a coherent and homogenous data object.	Req_005
DIS_12*	The BEYOND solution shall ensure that a mapping between data providers' ingested data to the common information model is established.	Req_006 Req_007
DIS_13*	The BEYOND solution shall allow data providers to edit and update the defined mapping.	Req_007
DIS_14*	The BEYOND solution shall allow data providers to select transformation rules (e.g., measurement units, measurement scale, time zones, etc) to be applied on their ingested data, and according to the common information model.	Req_008 Req_009
DIS_15*	The BEYOND solution shall enable semantic interlinking of different entities within the platform.	Req_010
DIS_16*	The BEYOND solution shall ensure that the quality of the ingested data are of high quality by combining them with other data on a semantic and payload level.	Req_010 Req_011
DIS_17*	The BEYOND solution shall support the inclusion of various metadata on the ingested data, to improve asset discoverability, and reveal relationships that might exist with other assets.	Req_012
DIS_18*	The BEYOND solution shall enable the fusion of different assets into a composite asset, to create new data profiles that carry more valuable information.	Req_012
DIS_19*	The BEYOND solution shall ensure that data transformations for measurement units are in accordance with the BEYOND CIM.	Req_008
PDSL_1	The BEYOND solution shall provide a consolidated storage component for the secure storage of all data assets ingested in the platform .	Req_042
PDSL_2*	The BEYOND solution shall enable data providers to store their data assets in a trusted data container, enforcing data security and privacy- ensuring mechanisms.	Req_042
PDSL_3*	The BEYOND solution shall enable storage of the data assets' metadata in a trusted data container.	Req_042
PDSL_4*	The BEYOND solution shall enable the storage of the access policies defined by the data providers on their assets.	Req_042

[*] Denotes that a particular functional requirement is also elicited for the Private Polyglot Data Storage Layer.

PDSL_5:	The BEYOND solution shall enable storage of the data analytics job related information's (e.g., trained algorithms, analytics models and results) in a trusted data container.	Req_042
PDSL_6*	The BEYOND solution shall enable the storage of all data ingestion job-specific configurations along with the transitional data files, enabling data providers if needed, to pause of a data ingestion job and undertake updates or resume it at a later stage.	N.A.
PDSL_7*	The BEYOND solution shall provide a secure storage container for saving all sensitive data and credentials, while enforcing controlled access over the different sensitive parameters, such as tokens, API keys, usernames, passwords, etc.	Req_022
PDSL_8	The BEYOND solution shall provide a secure storage container for saving all the various versions of the BEYOND CIM, inclusive of the model's concepts, fields and their relations.	Req_022
PDSL_9	The BEYOND solution shall provide a secure storage container for saving all the various versions of the domain vocabularies.	Req_022
PDSL_10	The BEYOND solution shall provide a dedicated distributed ledger for the secure storage of all contract-related data, while ensuring the privacy of the involved stakeholders.	Req_063
DEM_1	The BEYOND solution shall enable users (i.e., data asset consumers and/or data brokers) to search for data assets available in the platform.	Req_073
DEM_2	The BEYOND solution shall provide a suitable searching mechanism supporting both keyword and/or text-based queries from the users.	Req_079
DEM_3	The BEYOND solution shall enable users to save their search queries and update them if needed at a later stage.	Req_079
DEM_4	The BEYOND solution shall enable users to apply filters on their search queries, based on the data asset's information (e.g., relevant domain, data type, data format, language, etc.).	Req_079 Req_080
DEM_5	The BEYOND solution shall deliver an appropriate query definition mechanism providing a simple process for data asset searching (including AI models and analytics reports) available in the platform.	Req_079
DEM_6	The BEYOND solution shall provide a suitable searching mechanism enabling users to view the results (i.e., data assets) related to their search queries, along with the supplementary information; supporting an easier and faster identification.	Req_080
DEM_7	The BEYOND solution shall provide a suitable searching mechanism capable of providing recommendations of data assets or potential combinations of them, relevant to the users search queries, in case no relevant data asset is	Req_082

	available in the platform.	
DEM_8	The BEYOND solution shall provide a suitable searching mechanism capable of recommending to data consumers, potential BEYOND stakeholders that could provide and/or create a data asset, relevant to their search query, in case no relevant asset is available in the platform.	Req_082
DEM_9	The BEYOND solution shall enable users to view a sample of the recommended data assets regardless of the assets' access policies.	N.A.
DEM_10	The BEYOND solution shall enable users to request access to a data asset they are not eligible to view/acquire from the relevant data provider.	Req_044 Req_077
DEM_11	The BEYOND solution shall enable users to save the recommended data sets in their favourite lists, in order to be used again in the future for enhancing the recommendations results.	N.A.
DEM_12	The BEYOND solution shall enable users to view a list of all the recommendations for data assets and potential data providers, in case no asset relevant to their search query is available in the platform.	Req_080
IDAC_01*	The BEYOND solution shall provide data preparation functions (e.g., sorting, appending, etc) prior their propagation to the various algorithms.	Req_031
IDAC_02*	The BEYOND solution shall allow users to design and configure the execution of analytics processes on their data.	Req_030
IDAC_03*	The BEYOND solution shall allow users to decode encrypted data such that appropriate data preparation functions to be configured and executed.	Req_032
IDAC_04*	The BEYOND solution shall allow users to add import blocks to propagate input datasets to data analytics processes.	Req_032
IDAC_05*	The BEYOND solution shall allow users to add manipulation blocks to apply preparation functions prior to data analytics processes.	Req_033
IDAC_06*	The BEYOND solution shall allow users to add analysis blocks to apply different data analytics and machine learning techniques over the imported data.	Req_033
IDAC_07*	The BEYOND solution shall allow users to add export blocks to allow the generation of the analytics results.	Req_033

* Denotes that a particular functional requirement is also elicited for the Private Polyglot Data Storage Layer.

IDAC_08*	The BEYOND solution shall provide an adequate list of AI algorithms (e.g., regression, classification, clustering, etc) to be utilised and applied on users' data.	Req_032
IDAC_09*	The BEYOND solution shall allow users to configure the execution parameters of the available AI algorithms.	Req_033
IDAC_10*	The BEYOND solution shall provide pre-trained analytics models to be utilised for reduced analytics execution time.	Req_030
IDAC_11	The BEYOND solution shall execute all the data analytics processes in the BEYOND Cloud based Platform.	Req_034
IDAC_12*	The BEYOND solution shall enable users to select the format (e.g., csv, export for visualisation, etc) of the data analytics results.	N.A.
IDAC_13*	The BEYOND solution shall provide users a complete list of different visualisation graphs (e.g., line chart, bar charts, scatter plots, etc) for the visualisation of analytics' results.	Req_164, Req_165, Req_166, Req_167 Req_168, Req_169
IDAC_14*	The BEYOND solution shall allow users to configure the visualisation of the analytics' results.	Req_164, Req_165, Req_166, Req_167 Req_168, Req_169 Req_170
CPOM_01*	The BEYOND solution shall provide an appropriate authentication engine capable of generating and securely managing the identity information of users belonging to an organisation and which are authorised to access the system.	N.A.
CPOM_02*	The BEYOND solution shall provide an appropriate authentication engine for the management of the registration process of organisations and for the validating the information submitted regarding their legal entity and responsible persons (e.g., the responsible manager for signing the data assets' contracts).	N.A.
CPOM_03*	The BEYOND solution shall provide an appropriate authentication and authorisation service to the end-users tools for defining appropriate policies per tool, as well as generate and manage the relevant tokens.	Req_040
CPOM_04*	The BEYOND solution shall be able to manage the exchange of decryption keys, for the utilisation of encrypted data between the respective components of the platform ensuring end-to-end security.	Req_022

* Denotes that a particular functional requirement is also elicited for the Private Polyglot Data Storage Layer.

CPOM _05*	The BEYOND solution shall provide different mechanisms to verify the reliability and source of the data exchanged between the various components of the system.	Req_024
CPOM _06*	The BEYOND solution shall provide an appropriate notification mechanism able to identify the events taking place in the various components of the platform and their related recipients and generate the appropriate notifications to inform them.	Req_021
CPOM _07*	The BEYOND solution shall provide an appropriate notification mechanism able to generate and send appropriate notifications to the users regarding the execution progress of a data ingestion job (e.g., completion, failure, etc.).	Req_021
CPOM _08*	The BEYOND solution shall provide an appropriate notification mechanism able to generate and send appropriate notifications to the relevant users regarding the progress of a data asset brokerage activity (e.g., when it is due, when it is signed, rejected, etc.)	Req_021
CPOM _09*	The BEYOND solution shall provide an appropriate notification mechanism able to generate and send the appropriate notifications to the relevant users both through the platform and/or via email.	Req_021
CPOM _10*	The BEYOND solution shall enable users to view and manage all the notifications related to their organisation.	N.A.
CPOM _11*	The BEYOND solution shall enable users to set their preferences on how to receive notifications (through the platform or via email) and based on their content.	N.A.
CPOM _12*	The BEYOND solution shall be able to dynamically orchestrate the available resources for the successful execution of each data check -in and/or analysis job.	N.A.
CPOM _13	The BEYOND solution shall be able to dynamically set and assign the appropriate computing and storage resources required for the successful operation of the platform and the efficient execution of its services.	N.A.
CPOM _14	The BEYOND solution shall be able to create and manage API keys for the different end-users tools that have a retrieval scope and a specific expiry date.	Req_019
CPOM _15	The BEYOND solution shall be able to manage the different tasks included in receiving and processing simultaneous API calls, such as traffic management, version management, pagination and monitoring.	Req_019
CPOM _16	The BEYOND solution shall enable the platform administrators to monitor the platform's usage and services operations so as to plan maintenance activities, identify faults and timely carry out corrective actions.	N.A.

CPOM _17*	The BEYOND solution shall enable data asset providers to have an overview of their data assets' usage within the platform, as well as get an overview of all their data ingestion and analytics jobs.	
CPOM _18*	The BEYOND solution shall ensure that only anonymised data are collected towards extracting insights on the platform's usage, always adhering to the respective GDPR regulations.	Req_023
DTM_01	The BEYOND solution shall provide a trusted data sharing mechanism utilising a blockchain based data marketplace.	N.A.
DTM _02	The BEYOND solution shall provide an appropriate interface where data consumers/data brokers will be able to enter into multi-party contract with data owners for the acquisition of their data assets.	Req_047
DTM _03	The BEYOND solution shall provide an appropriate mechanism enabling users to draft smart asset contracts for the acquisition of data assets	Req_049
DTM _04	The BEYOND solution shall enable data owners to define their preferred license, IPR details and pricing scheme for their data assets.	Req_047
DTM _05	The BEYOND solution shall provide an appropriate mechanism enabling users to negotiate the terms and respective fee of data asset contract.	Req_050, Req_052
DTM _06	The BEYOND solution shall provide an appropriate mechanism for evaluating any attempted data asset activity against the respective contract's terms to ensure that the terms are not violated.	Req_057
DTM _07	The BEYOND solution shall provide an appropriate mechanism for checking if a smart data contract is active and valid, prior to enabling access to the data consumer/broker.	Req_061
DTM _08	The BEYOND solution shall enable data consumer/brokers to create their own wallet to add funds, see their transaction history and pay the respective fees of a smart data asset contract.	Req_058, Req_065
DTM _09	The BEYOND solution shall provide an appropriate remuneration mechanism, capable of managing the payments related to the activation of a smart contract.	Req_056
DTM_10	The BEYOND solution shall enable the parties involved in a smart asset contract to download the contract terms (e.g., pdf) for evaluation and archiving purposes.	Req_064
DTM_11	The BEYOND solution shall utilise a trusted cryptocurrency technology for the settlement of the smart data asset contracts	Req_058
DTM_12	The BEYOND solution shall enable data owners to create their own wallet, see their transaction history and receive funds related to the activation of a smart asset contract.	Req_058, Req_065

DTM_13	The BEYOND solution shall enable also conventional payments (e.g., bank transfer) to the account of the data provider, or via the credit card of the data consumer.	Req_058

6.2 List of BEYOND Non-Functional requirements

Having described above the main functional requirements from the BEYOND Platform components' perspective, Table 6-2 below presents a non-exhaustive list of non-functional requirements for the BEYOND platform.

TABLE 6-2 BEYOND PLATFORM - NON-FUNCTIONAL REQUIREMENTS

ID	Description
NFR_01	The platform shall provide an intuitive and user-friendly interface to users.
NFR_02	The platform shall allow for user registration and login with credentials.
NFR_03	The platform shall allow only secure communication.
NFR_04	The platform shall be accessible from a browser
NFR_05	The platform shall offer easy navigation
NFR_06	The End-User Tools shall be provided in English language
NFR_07	Each End-User Tool shall be provided in the official language of its corresponding organisation
NFR_08	The platform shall be provided in English language
NFR_09	The platform shall be provided in all demo sites' languages
NFR_10	The platform should keep global behaviour consistency
NFR_11	The platform should be capable to maintain cross-browser functionality
NFR_12	The platform should process all requests that are not related to data analytics and data operations within 5 seconds
NFR_13	The platform should be capable to respond to requests within 1 second
NFR_14	The platform should be interoperable such that its services are interfaced with other systems without restrictions

7 BEYOND End-Users Tools – Detailed Architecture

The BEYOND End-Users Tools aim to cope with certain energy and building challenges identified by the different stakeholders through the utilisation of different functionalities offered by the BEYOND Platform. The platform's core functionalities that facilitate the use of end-users tools include the data ingestion, data exploration, and data analysis. In particular, an end-user tool may provide datasets extracted from their applications to the BEYOND Platform following the data ingestion process (see Section 1 for a detailed description on data ingestion).

Furthermore, an end-user tool can retrieve data (see Section 5.2.2 for a detailed description on data exploration) already loaded in the BEYOND Platform, through API requests and based on the signed contracted between the end-user tool and the data provider that offered the data to the platform.

Finally, end-user tools can exploit the different analytics processes that are available through the data analysis functionalities, by designing and applying data analytics workflows in any of the two layers provided by the BEYOND Platform (i.e., BEYOND Cloud based Platform or BEYOND Private Infrastructure).

Figure 7-1 depicts the list of all end-users tools provided by the BEYOND business developers. A detailed description including a brief overview, the functional and non-functional requirements, the architecture and components, as well as the process view for each of the end-users tools are provided accordingly in the subsequent subsections aiming to describe the different tools that will be provided by WP5 and WP6 of the project.

FIGURE 7-1: LIST OF END-USERS TOOLS

The different requirements from each tools are also linked to the use cases that have been identified in deliverable D2.1.

7.1 Impact Assessment Tool for Energy Policy Making at Urban Level (EPUL)

7.1.1 Brief Description

The aim of the EPUL Tool (stepping on ARTELYS Crystal City) is to help local authorities to elaborate their Sustainable Energy and Climate Action Plan (SECAP) and make investment decisions that will enable them to meet the European Union's commitments in terms of reducing GHG emissions at a lower cost while improving the quality of life of the inhabitants.

This tool will provide local authorities with an accurate picture of the energy consumption of their territory, of the equipment (type of boilers, hot water, etc.) and of the uses of the inhabitants (particularly in terms of transport) in order to identify the levers for action.

Indeed, this mapping of the territory and its analysis is essential to any bottom-up approach to assessing the impact of investment decisions on the GHG emissions trajectory.

The forecasting capabilities and simulation accuracy of the Crystal City Tool will be enhanced by the baseline industrial data analytics developed in WP4 in order to provide better informed evidence to local decision makers (urban level) for optimized energy policy making.

Instead of forecasting the future energy performance of buildings (which constitute the major energy demand in urban contexts) based on high-level and low granularity metering data, the EPUL tool (the Beyond extended version of Crystal City tool) will be equipped with building demand forecasts of higher accuracy to enhance its predictive capabilities and, thus, allow the tool to provide a more accurate representation of the environmental and energy state in specific urban contexts, subsequently enabling the better informed identification of local energy/sustainability requirements and the design of more realistic policy measures to achieve mid- and long-term sustainability objectives.

Demand (and generation) forecasting models for different types of buildings (resulting from T4.2 and utilizing highly granular and diverse building data to achieve high accuracy) will be made available to the EPUL tool, to enable a more representative and realistic population of the tool's impact assessment (simulation) models and, subsequently, more accurate and robust predictions of the impact achieved by alternative policy scenarios designed by city authorities with the ultimate objective to support them in making the optimal decision for the timely and effective satisfaction of energy and sustainability objectives set out in local authorities' action plans for the mid- and long-term.

7.1.2 Functional and Non-Functional Requirements

The realization of the functionalities of the Impact Assessment Tool for Energy Policy Making at Urban Level comes with a list of functional requirements as listed in the table below:

TABLE 7-1 EPUL FUNCTIONAL REQUIREMENTS

Functional Requirement ID	Description	Relevant Use Case(s)
EPUL_001	The EPUL Tool should allow to design policy scenarios.	UC_5.7 UC_5.8
EPUL_002	The EPUL Tool should allow to define parameters of policy scenarios.	UC_5.7 UC_5.8
EPUL_003	The EPUL Tool should allow to the user to populate the assets involved in each scenario	UC_5.7 UC_5.8
EPUL_004	The EPUL Tool should provide highly accurate building demand forecasts	UC_5.7 UC_5.8
EPUL_005	The EPUL Tool should provide a more accurate representation of the environmental and energy state in specific urban contexts	UC_5.7 UC_5.8
EPUL_006	The EPUL Tool should enable a well-informed identification of local energy / sustainability requirements	UC_5.7 UC_5.8
EPUL_007	The EPUL Tool should enable the design of more realistic policy measures to achieve mid- and long-term sustainability objectives	UC_5.7 UC_5.8
EPUL_008	The EPUL tool should have access to demand/generation forecasting models for different kinds of buildings	UC_5.7 UC_5.8
EPUL_009	The EPUL tool should have access to batch static data from well-known repositories of open building data	UC_5.7 UC_5.8
EPUL_010	The EPUL Tool should allow to the user to visualize the results of simulations and a variety of KPIs relevant to SECAP	UC_5.7 UC_5.8

EPUL_011	The EPUL Tool should allow to the user to compare the results of the simulated scenarios two by two	UC_5.7 UC_5.8
EPUL_012	The EPUL Tool should allow to the user to identify weaknesses and re-configure an existing scenario	UC_5.7 UC_5.8

In addition to the functional requirements presented above, a list of non-functional requirements has been elicited and presented in the following table:

TABLE 7- 1 EPUL NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement ID	Description
EPUL_N_001	The tool shall be privacy aware (complaining with pseudonymised, anonymised as well as personal data principles)
EPUL_N_002	The tool shall be always up to date with the latest data available
EPUL_N_003	The tool shall have login requirements
EPUL_N_004	The tool should be enterable with open standards
EPUL_N_005	The tool shall interact with BEYOND platform (data access & sharing)
EPUL_N_006	The tool shall provide an intuitive and user-friendly interface to the user
EPUL_N_007	The tool shell be documented
EPUL_N_008	The tool shall have sufficient performance in loading and update times
EPUL_N_009	The tool shall have high performance in the presentation of simulation results
EPUL_N_010	The tool should be compatible with design standards
EPUL_N_011	The tool shall be reliable, scalable and available

7.1.3 Architecture - Components View

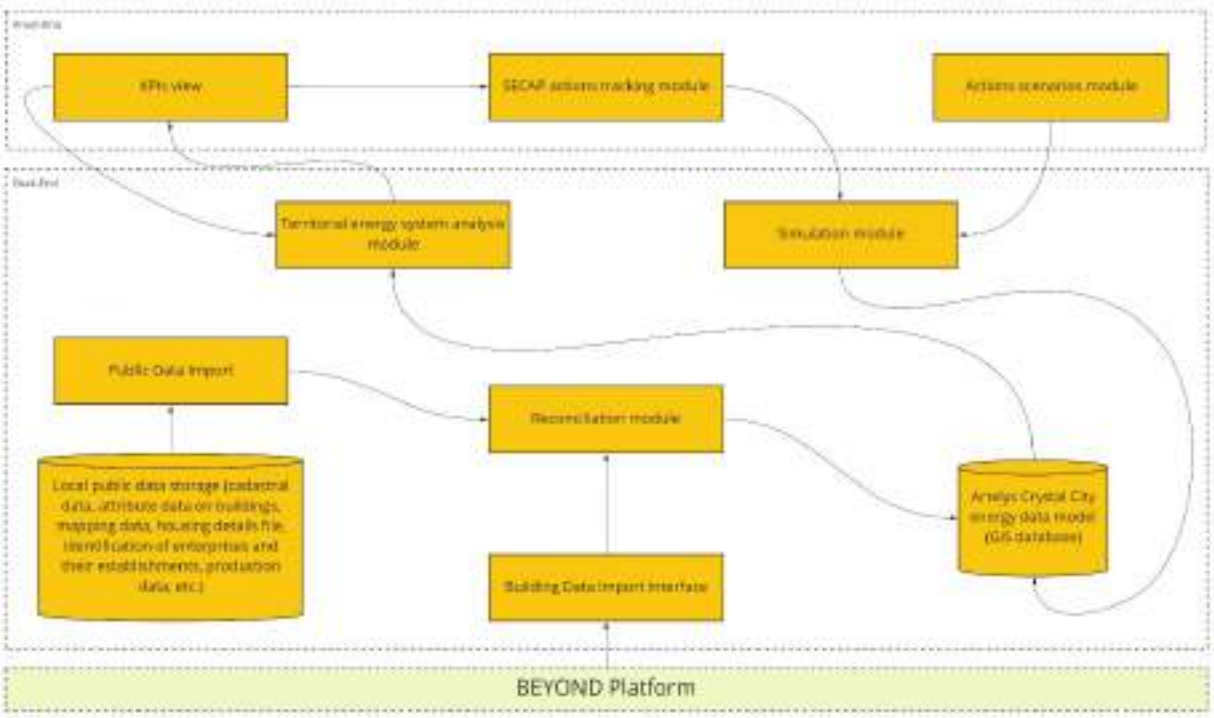

FIGURE 7-2: EPUL ARCHITECTURE

EPUL-1: Local public data storage: this database is intended to host public data useful for mapping the energy consumption of the territory (cadastral data, attribute data on buildings, mapping data, housing details file, identification of enterprises and their establishments, production data, etc.). These data are generally accessible at a coarser scale than the building (district, municipality, NUTS3).

EPUL-2: Public Data Import: the purpose of this module is to parse and import public data which will then be processed by the reconciliation module.

EPUL-3: Building Data Import Interface: this module is designed to parse and import data at the building level stored by the BEYOND platform which will then be processed by the reconciliation module. At this stage, processing will be carried out to remove outliers, interpolate missing data, and then aggregate them to the target grid (district, municipality, NUTS3, etc.) of the SECAP study.

EPUL-4: Reconciliation module: the reconciliation module aims at reconciling the different sources of data on different levels of aggregation to constitute the energy model of the reference territory stored in a pivot format.

EPUL-5: Artelys Crystal City energy data model: this database contains a mapping of the territory's energy system (consumption broken down by consumer category (typology), supply energy and usage, production infrastructure and networks) in the Artelys Crystal City pivot format.

EPUL-6: Territorial energy system analysis module: computation engine of parameterizable indicators allowing the analysis of the energy options of the territory.

EPUL-7: KPIs view: this module allows the user to set the KPIs that seem relevant to evaluate his territorial energy system and to visualize the results.

EPUL-8: Actions scenarios module: this module allows the user to define mid or long-term scenarios based on the combination of easily parameterizable unit actions.

EPUL-9: Simulation module: computation engine of techno-economic simulation of system operation.

EPUL-10: SECAP actions tracking module: this module will allow the user to monitor the implementation of the actions included in the SECAP, to measure their real impact (difference between the simulation results and the observed emissions) and to propose amendments in order to remain on track.

7.1.4 Sequence Diagram

The flow diagram below presents the interaction flow between the components represented in the architecture picture earlier.

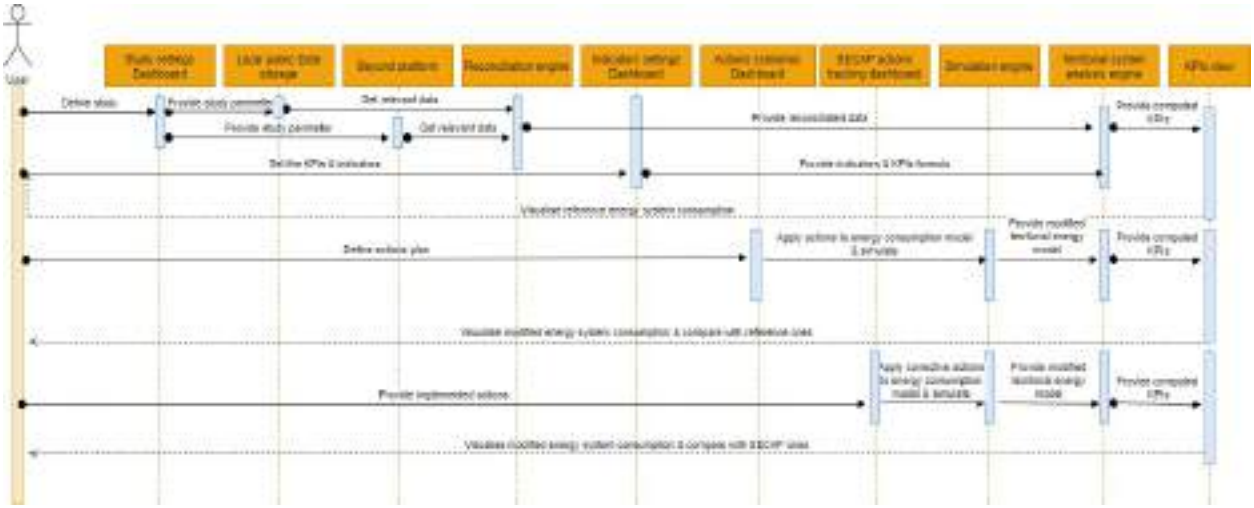

FIGURE 7-3: EPUL FLOW DIAGRAM

The GUI allows user to input perimeter of the study (study area, granularity, time horizon, etc.). Based on that information, relevant data will be extracted from public local data storage and Beyond platform. These data will be reconciled at the same level of granularity.

On the basis of these data, reference territorial energy model will be set up. The user will also be able to configure the indicators and KPIs to be calculated. After launching the simulation, he will be able to visualize the result of these indicators on his reference territorial energy model.

Then, the user will define, through the GUI, a scenario as a combination of unitary energy actions. These actions will be applied to the territorial energy reference model which will be distorted. After launching the simulation, the user will be able to visualize the results [of his indicators] and compare them with the results of the reference territorial energy model.

Once the user has refined his scenario, he can define it as a SECAP scenario. During the implementation of SECAP, the user will be able to inform the progress of the actions. In case of deviation from the objectives, the tool will propose corrective actions to be implemented which can then be simulated. The results can then be visualized and compared with the initial plan (the aim will be to get as close as possible).

7.2 District Heating Network Planning and Infrastructure Sizing Tool (DHCO)

7.2.1 Brief Description

The District Heating & Cooling Optimizer (DHCO) is an innovative tool that will utilize input (policy making optimization targets and needs) coming from the Impact Assessment toolbox (T5.1) to provide decision-makers and project developers with district heating and cooling infrastructure sizing and expansion planning to satisfy future needs of energy networks and infrastructure in response to specific energy policy objectives.

The DHCO Tool will simulate the operation of the networks in appropriate horizons (related to regulatory regimes and asset lifetimes) and calculate energy network reliability, performance and quality metrics based on different infrastructure and assets installation setups, their characteristics and planned/unplanned events.

The algorithms will incorporate functionality to assess the performance of the networks under various demand uncertainties. They will be appropriately configured to enable the study of optimal planning and sizing of energy networks assets such as new connections, pointing out to specific reinforcement/ expansion needs to accommodate large volumes of new connections and assets or the degree of flexibility required, considering also asset management input and needs for further investments.

The algorithms will have the capability to embed all information arising from Network Operator systems, flexibility analytics, short-, mid- and long-term forecasting analytics, along with batch static data coming from well-known hubs of open building data in order to perform a comprehensive simulation-based analysis of performance and reliability metrics in specific parts of district heating networks under different alternative penetration scenarios for new energy infrastructure and assets.

7.2.2 Functional and Non-Functional Requirements

The realization of the functionalities of the District Heating Network Planning and Infrastructure Sizing Tool comes with a list of functional requirements as listed in the table below:

TABLE 7-2 DHCO FUNCTIONAL REQUIREMENTS

Functional Requirement ID	Description	Relevant Use Case(s)
DHCO_001	The DHCO should provide highly accurate building demand forecasts	UC_5.4 UC_5.5 UC_5.6
DHCO_002	The DHCO should enable clustering of buildings	UC_5.4

	based on their demand characteristics, typology, occupants' synthesis and other parameters for which data exist	UC_5.5 UC_5.6
DHCO_003	The DHCO should enable the creation of a scenario (building stock expansion, adjustment of modelling artefacts/ buildings to accurate demands, extrapolation of selected building clusters) defining study area, time horizon and connection target	UC_5.4 UC_5.5 UC_5.6
DHCO_004	The DHCO should compute optimal network layout (in the economic sense) and sizing of its components (pipes, etc) based on accurate building demand forecasts	UC_5.4 UC_5.5 UC_5.6
DHCO_005	The DHCO should select the buildings to be connected to the network	UC_5.4 UC_5.5 UC_5.6
DHCO_006	The DHCO should allow, via the cartographic module, to exclude manually (or via the application of filters) roads and buildings from the route	UC_5.4 UC_5.5 UC_5.6
DHCO_007	The DHCO should select the most relevant production sources to ensure a balance in supply and demand	UC_5.4 UC_5.5 UC_5.6
DHCO_008	The DHCO should provide indicators on the need for production capacity	UC_5.4 UC_5.5 UC_5.6
DHCO_009	The DHCO should display optimal network on a map	UC_5.4 UC_5.5 UC_5.6
DHCO_010	The DHCO should compute and display KPIs: cost, compliance with physical constraints, etc.	UC_5.4 UC_5.5 UC_5.6

In addition to the functional requirements presented above, a list of non-functional requirements has been elicited and presented in the following table:

TABLE 7-3 DHCO NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement ID	Description
DHCO_N_001	The tool shall be privacy aware (complaining with pseudonymised, anonymised as well as personal data principles)
DHCO_N_002	The tool shall be always up to date with the latest data available
DHCO_N_003	The tool shall have login requirements
DHCO_N_004	The tool should be enterable with open standards
DHCO_N_005	The tool shall interact with BEYOND platform (data access & sharing)
DHCO_N_006	The tool shall provide an intuitive and user-friendly interface to the user
DHCO_N_007	The tool shell be documented
DHCO_N_008	The tool shall have sufficient performance in loading and update times
DHCO_N_009	The tool shall have high performance in the presentation of simulation results
DHCO_N_010	The tool should be compatible with design standards
DHCO_N_011	The tool shall be reliable, scalable and available

7.2.3 Architecture - Components View

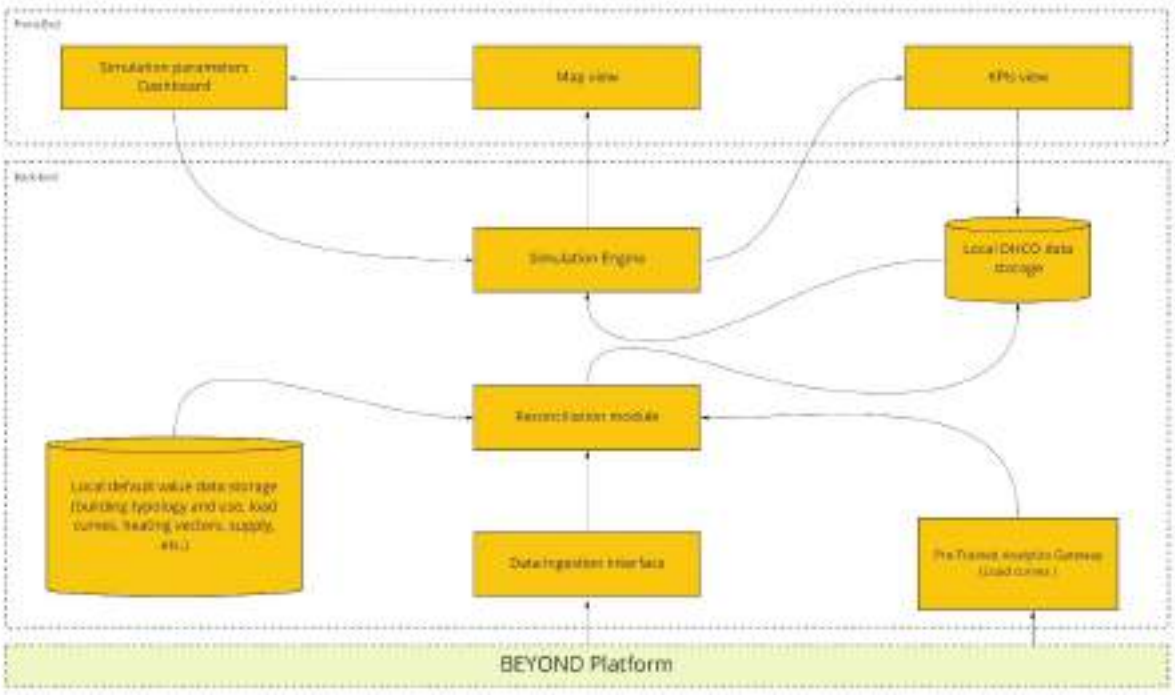

FIGURE 7-4: DHCO ARCHITECTURE

DHCO-1: Local default value data storage: this database is intended to host default value data useful for interpolating missing information and checking the consistency of the data provided.

DHCO-2: Data Ingestion Interface: the purpose of this module is to parse and import building data provided by Beyond platform that will then be processed by the reconciliation module.

DHCO-3: Pre-Trained Analytics Gateway: the objective of this module is to calculate the load curves of buildings based on their typology, use and living area. A consumption forecasting model will be used to adjust the load curve over the simulated time horizon.

DHCO-4: Reconciliation module: the reconciliation module aims at reconciling the different sources of data on different levels of aggregation to constitute the relevant data model that will be use to size the heating and cooling network.

DHCO-5: Local DHCO data storage: this database contains a mapping of the territory's energy system (load curve and heating vector of each building, typology of each building, production capacity, etc.) in the Artelys Crystal City pivot format.

DHCO-6: Simulation parameters Dashboard: dashboard that allows the user to parameterize his simulation (time horizon, connecting goal, exclusions, risk policy, etc.).

DHCO-7: KPIs view: this module allows the user to set the KPIs that seem relevant to evaluate his study and to visualize the results.

DHCO-8: Simulation engine: technical and economic optimization engine which takes into account the investment cost and the operational cost.

DHCO-9: Map view: this module will allow the user to visualize the resulting network routes.

7.2.4 Flowchart Diagram

The flow diagram below presents the interaction flow between the components represented in the architecture picture earlier.

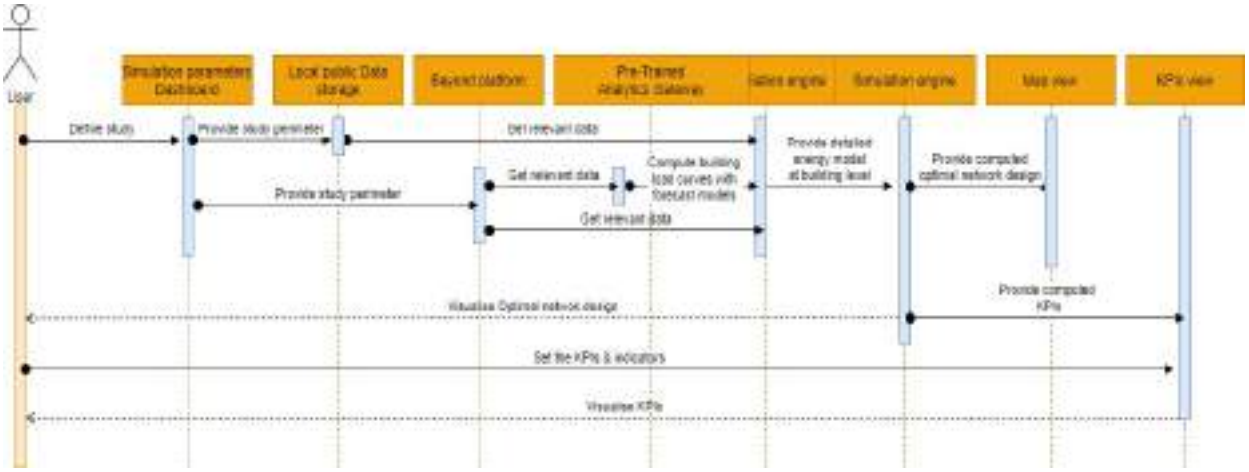

FIGURE 7-5: DHCO FLOW DIAGRAM

The GUI allows user to input perimeter of the study (study area, time horizon, targets, etc.). Based on that information, relevant data will be extracted from public local data storage and Beyond platform. These data will be pre-processed using a forecasting module and then reconciled in order to obtain a forecasted load curve per building.

The simulation engine will calculate an optimal network layout (in the economic sense) and select the buildings to be connected.

The user will also have the possibility, via the cartographic module, to exclude manually (or via the application of filters) roads and buildings from the route.

The simulation engine will also select the most relevant production sources to ensure a balance in supply and demand. Indicators on the need for production capacity will also be offered to the user.

The optimal layout of the network will then be displayed within the cartographic module and indicators (cost, compliance with physical constraints, etc.) will be proposed to the user.

7.3 Distribution Grid Planning and Infrastructure Sizing Tool (DGPIST)

7.3.1 Brief Description

The main function of the Distribution Network Infrastructure and Planning Tool is to obtain the best techno-economic planning by optimally managing network assets and the available flexibility.

The tool evaluates the grid state for different scenarios of generation and consumption according to different time horizons (short and long term). When generation or demand growth is necessary performed investments of new network infrastructure. The network operators want to reduce the total cost on network planning while the network keeps working properly, so this tool aim is to show several solutions and point out the best alternative.

To manage the network properly, it is necessary to know in depth the connected elements characteristics and their behaviour. Customer clustering analyses all customer consumption and classify them into types and evaluates the flexibility of each cluster. When a new customer is connected to the network, there are no information about its consumption and flexibility. The clustering allows include the new customer into one of the clusters according to the information available as maximum power or type (residential, industrial...) and, in this way, the tool assumes that new users can offer their cluster flexibility.

Network weaknesses must be analysed for different time horizons according to expected network expansion. As a result, weaknesses of the grid are pointed out and several solutions are analysed to improve the grid behaviour. These solutions include available flexibility management and, when this is not enough, the necessary investments in upgrade the network to increase its capacity.

The Distribution Grid Planning and Infrastructure Sizing Tool is useful to obtain an optimal planning. With the tool functionalities, network operators can evaluate the actual and the future network needs (by identifying weaknesses) and the best technical and economical option to solve operational problems.

7.3.2 Functional and Non-Functional Requirements

Next table shows the Distribution Network Infrastructure and Planning Tool functional requirement.

TABLE 7-4 DGPIST FUNCTIONAL REQUIREMENTS

Functional Requirement ID	Description	Relevant Use Case(s)
DGPIST_001	The Distribution Grid tool shall forecast generation and consumption in different temporal granularity: long	UC_5.3, UC_5.4,

	term to analyse future scenarios of the network and short term to manage the network operation	UC_5.6
DGPIST_002	The Distribution Grid tool shall enable the creation of a complete network model on PowerFactory with different scenarios of generation and demand to simulate the base case (only generation and demand forecasting) and different network expansion scenarios (different demand or generation growths)	UC_5.3, UC_5.4
DGPIST_003	The Distribution Grid Tool shall analyse the results obtained on the simulations and calculates several KPIs	UC_5.3, UC_5.4
DGPIST_004	The Distribution Grid Tool shall identify network weaknesses)e.g. congestions or voltage violations) by comparing the KPIs obtained with the expected KPI value in normal operation.	UC_5.3, UC_5.4
	The Distribution Grid tool shall provide a visualisation of the KPIs obtained and the expected values.	UC_5.3, UC_5.4
DGPIST_005	The Distribution Grid Tool shall enable the visualisation of a list of elements of the grid where a weakness has been detected and a list of assets reinforcements or flexibility needs to remove the grid weakness.	UC_5.3, UC_5.4
DGPIST_006	The Distribution Grid Tool should implement network upgrades according to the solutions proposed on a simulation and recommend the most optimal one from a technical point of view.	UC_5.4
DGPIST_007	The Distribution Grid Tool shall compare the KPIs obtained in different simulated scenarios (base case, network expansion, network upgrade…) to provide an in-depth analysis of the improvement on the grid behavior.	UC_5.4
DGPIST_008	The Distribution Grid Tool shall analyse the information available of all the network customers and classifies them in different groups in order to optimise the flexibility analysis (optimal number of groups).	UC_5.5
DGPIST_009	The Distribution Grid Tool shall identify the flexibility of each group of customers (customer clustering) at different temporal and spatial granularity in order to be able to manage it when network operations require it.	UC_5.5
DGPIST_010	The Distribution Grid Tool should analyse the information of a new customer and classify it into one of the previously defined groups based on this information (clusters).	UC_5.5
DGPIST_011	The Distribution Grid Tool shall analyse the characteristics of the flexibility offered by each customer in order to make a short-term forecast.	UC_5.5

DGPIST_012	The Distribution Grid Tool shall automatically manage the flexibility of the network to ensure the optimal network performance in different scenarios.	UC_5.6
DGPIST_013	The Distribution Grid Tool shall implement in the same simulation all the network upgrades proposed and the available flexibility for each time horizon.	UC_5.6
DGPIST_014	The Distribution Grid Tool shall analyse the simulation results and recommend the operational planning and flexibility needs to obtain robust network for different growth in demand and generation.	UC_5.6

In addition to the functional requirements presented above, a list of non-functional requirements has been elicited and presented in the following table:

TABLE 7-5 DGPIST NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement ID	Description
DGPIST_N_001	The application shall interconnect with the BEYOND Core Platform (data access & sharing).
DGPIST_N_002	The tool shall provide an intuitive and user-friendly interface to the user.
DGPIST_N_003	The tool shall have a high performance in the display of simulation results.
DGPIST_N_004	The tool shall be reliable and scalable
DGPIST_N_005	The tool should always be updated with the latest available data.
DGPIST_N_005	The application shall allow access to registered users (with credentials).
DGPIST_N_006	The tool shell be documented

7.3.3 Architecture - Components View

- **DGPIST-1: Planification Analytics Dashboard,** which enables the creation of dashboards for visualizing the results of the flexibility-based operational planning performed.
- **DGPIST-2: Customer Clustering Dashboard,** which enables the creation of dashboards for visualizing the results of the customer clustering performed.
- **DGPIST-3: Customer clustering engine,** this engine use machine learning technologies to classify the network customers into several representative types according to their energy consumption and the flexibility that they can offer.
- **DGPIST-4: Network management engine,** this engine manages the available flexibility of the network in each analysed time interval to ensure optimal network operation.

- **DGPIST-5: Infrastructure sizing engine,** this engine evaluates the network simulation results and the proposed solutions of network upgrades to remove the network weakness.
- **DGPIST-6: Network analysis engine,** this engine analyses the network state according to relevant KPIs and detect network weakness.
- **DGPIST-7: Network model simulation engine,** this engine model the network to be analysed in the software PowerFatory DIgSILENT and simulates several scenarios at different time horizons.

- **DGPIST-8: The pre-trained Analytics Gateway (Energy Consumption/Generation Forecasting)** is fed with data and analytics from the BEYOND cloud-based platform regarding consumption and generation forecasting and flexibility profiles.
- **DGPIST-9: The Data Ingestion Interface** is used to retrieve from the BEYOND Cloud-based platform data as customer information or network data.

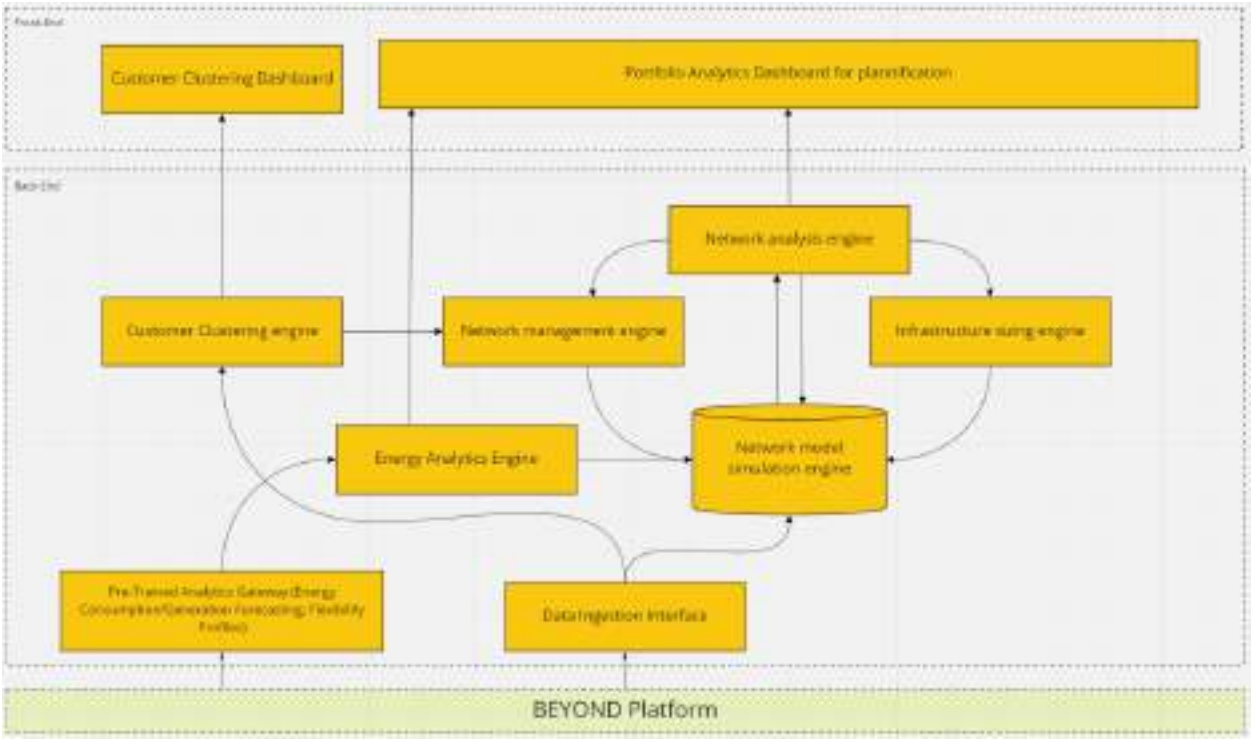

FIGURE 7-6: DGPIST ARCHITECTURE

7.3.4 Sequence Diagram

The following figure shows the flow diagram of how a user operates with the DPGIS tool.

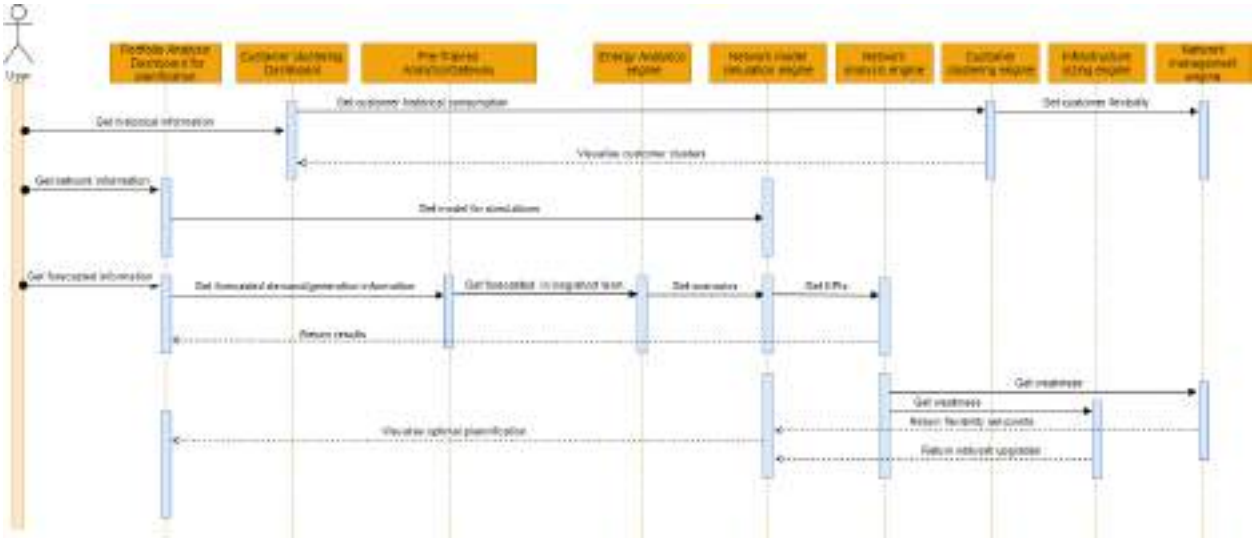

FIGURE 7-7: DGPIST FLOW DIAGRAM

As shown in the sequence diagram, there are two different workflows: customer clustering and network analysis. Customer clustering gets all the information regarding customers demand (consumption and flexibility) and by the comparison between them, the Customer Clustering Engine return all the customer cluster and the flexibility of each cluster.

The other workflow (network analysis) is also divided into the optimal management of the network assets flexibility tool and the new infrastructure sizing tool. Analysis Dashboard for planification can get all the information regarding network information and demand and generation forecasting by getting the data stored in the BEYOND Cloud based Platform. With these data, the network model is upgraded, and the Network analysis engine evaluates all the solutions proposed by the Network management engine and the infrastructure sizing engine.

Finally, the customer clusters from the first analysis and a list of the most suitable options to improve the network status to optimise the network planning are visualised.

7.4 Renovation Optimisation Decision Support Tool (ROST)

7.4.1 Brief Description

The aim of Renovation Optimisation Support Tools (ROST) is to facilitate the accurate energy-efficient design of buildings towards optimised investment decision-making when renovating the building.

The respective tool will utilise occupants' behaviour and comfort profiles resulting from WP4 and introduce them in iterative simulation loops of alternative renovation scenarios of selected buildings, thus replacing generic routines and schedules currently used in similar commercial products with real data coming from the actual operation of the building to- be-renovated.

Additional analytics will be performed over simulation results to identify energy performance de-efficiencies and enable further optimisation of renovation approaches and scenarios to achieve optimal balancing between anticipated energy performance, renovation project costs and air quality and occupants' comfort.

7.4.2 Functional and Non-Functional Requirements

The realisation of the functionalities of the Renovation Optimisation Decision Support Tool comes with a list of functional requirements as listed in the table below:

TABLE 7-6 ROST FUNCTIONAL REQUIREMENTS

Functional Requirement ID	Description	Relevant Use Case(s)
ROST_001	The tool shall provide a GUI for designing alternative renovation scenarios	UC_5.1 UC_5.2
ROST_002	The renovation too should assess real building energy performance (kWh/m2/year, CO2)	UC_5.1 UC_5.2
ROST_003	The renovation tool shall provide a visualisation / a list of building energy renovation solutions in use/actions made (e.g., windows change, energy assets replacements, etc.)	UC_5.1 UC_5.2
ROST _004	The renovation tool shall simulate different renovation scenarios against different parameters such as energy performance improvement needs, occupants' comfort profiles, budget, payback time, etc.	UC_5.1 UC_5.2
ROST _005	The renovation tool shall provide a visualisation of ranked (cost-optimised) list of renovation actions (achieved energy savings, payback time, CO2)). This will support identification of weak performance points.	UC_5.1 UC_5.2

ROST _006	The renovation tool shall provide accurate energy performance prediction (kWh/m2/year) of building generated by the selected renovation at the design time	
ROST _007	The Renovation Tool shall provide a comparative visualisation of real & simulated building energy performance (kWh/m2/year, CO2) of buildings before and after the implementation of renovation actions (using IPVMP approach).	UC_5.1 UC_5.2
ROST_008	Renovation tool output data should be accessible/exportable to Building Energy Passport tools	UC_5.1 UC_5.2

In addition to the functional requirements presented above, a list of non-functional requirements has been elicited and presented in the following table:

TABLE 7-7 ROST NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement ID	Description
ROST_N_001	The tool shall be privacy aware (complaining with pseudonymised, anonymised as well as personal data principles)
ROST_N_002	The tool shall be always up to date with the latest data available
ROST_N_003	The tool shall have login requirements
ROST_N_004	The tool should be enterable with open standards
ROST_N_005	The tool shall interact with BEYOND platform (data access & sharing)
ROST_N_006	The tool shall provide an intuitive and user-friendly interface to the user
ROST_N_007	The tool shell be documented
ROST_N_008	The tool shall have sufficient performance in loading and update times
ROST_N_009	The tool shall have high performance in the presentation of simulation results
ROST_N_010	The tool should be compatible with design standards
ROST_N_011	The tool shall be reliable, scalable and available

7.4.3 Architecture - Components View

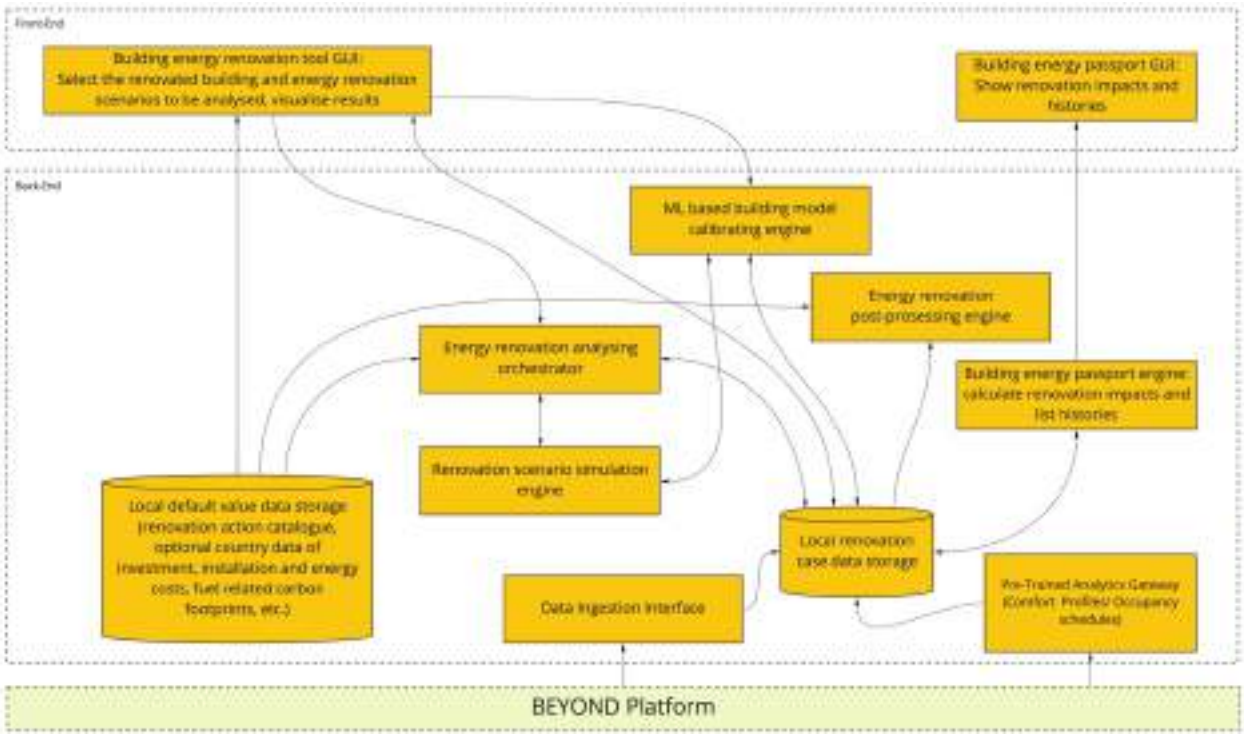

FIGURE 7-8: ROST ARCHITECTURE

ROST-1: Building energy renovation tool GUI: This GUI makes it possible to read the building basic values (e.g., building height, floor area, etc.) from local database (optionally communicating with BEYOND platform in the background) or input related values manually.

ROST-2: ML based building model calibrating engine: This software calibrates the building model parameters by minimizing the error with the building energy model and related measurement (heating, cooling, electricity and optionally water consumption, local weather data). The minimizing will be done by machine learning (ML) technologies.

ROST-3: Energy renovation analysing orchestrator: This software call in loop (all user selected renovation scenarios) the renovation scenario engine, reads needed inputs and save results in local renovation case data storage.

ROST-4: Renovation scenario simulation engine: This calculation engine simulates the given building energy renovation case related energy consumption.

ROST-5: Energy renovation post-processing engine: This software post-process calculated renovation case information. This includes renovation cost and carbon footprint calculation (if country level information is available) and visual based prioritizing the calculated renovation scenarios.

ROST-6: Local renovation case data storage: This database includes needed renovation software input and output values (optionally communicating with BEYOND platform in the background).

ROST-7: Local default value database: This database includes renovation related background information like renovation data, country level energy, renovation work and investment costs, used fuel related carbon footprints and selected data for typical values for some building types. At this moment these values are available only for Finland building stock.

ROST-8: Building energy passport GUI: This GUI makes it possible to list the renovation histories and visualise related renovation impacts.

ROST-9: Building energy passport engine: This software reads the renovation histories from the local database (optionally communicating with BEYOND platform in the background) and calculating studied renovation actions impacts (e.g., energy savings).

7.4.4 Sequence Diagram

The flow diagram below presents the interaction flow between the components represented in the architecture picture earlier.

The GUI allows user to input building basic values. It can also start machine learning based calibration the building energy model parameters, so that the real consumption of the building and building energy model give almost the same values. In addition, the user can select interested renovation scenarios from local default value database (includes renovation action catalogue), start selected renovation actions analyses and visualise related results.

All information inserted by the user or accessed through the platform is stored in the local ROST database (DB).

The results of renovation simulations in form of ranked renovation actions are returned to the GUI to be visualised to the user of the tool.

Further information of performed renovation actions are exported to support building energy passport service.

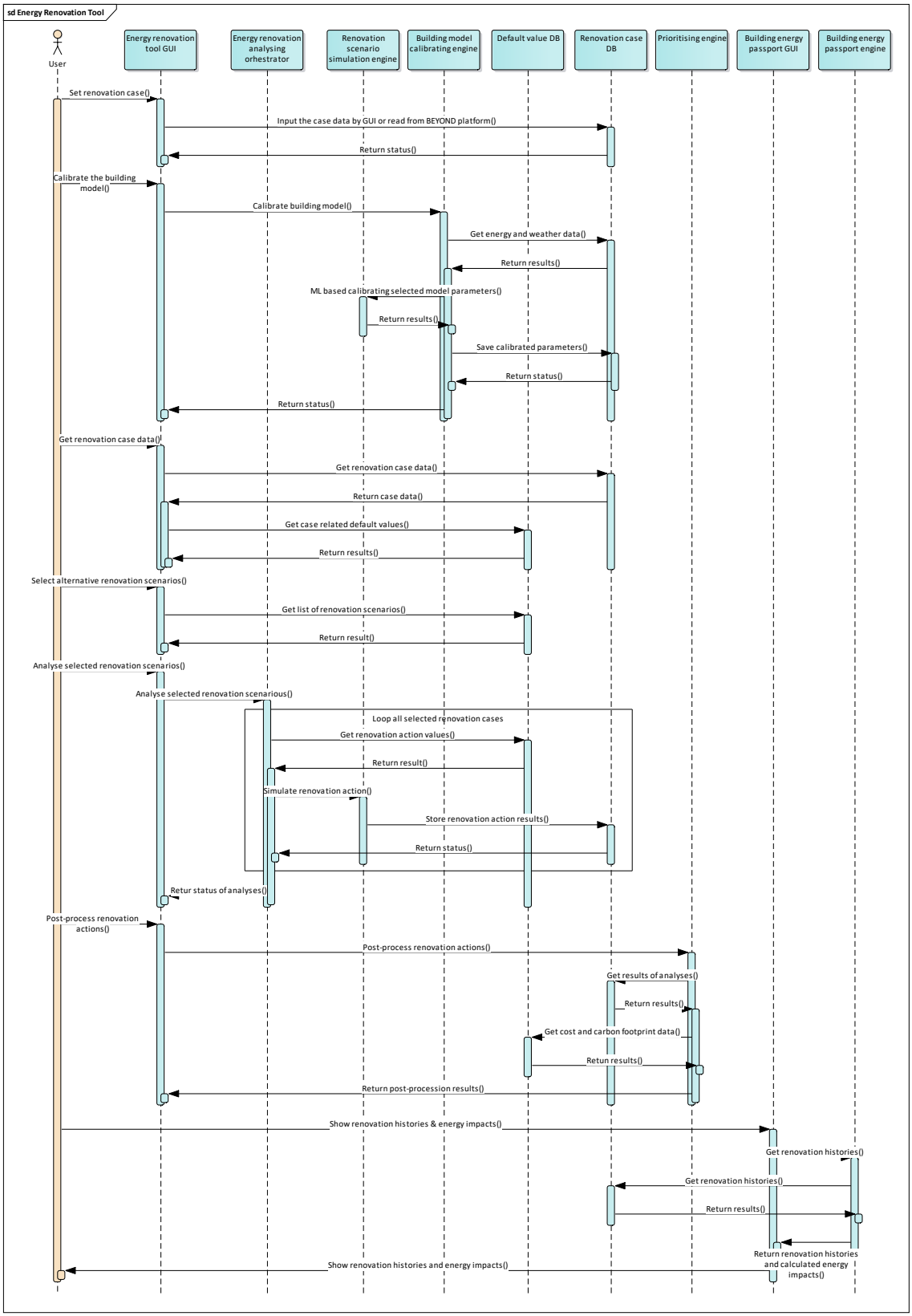

FIGURE 7-9: ROST FLOW DIAGRAM

7.5 Building Digital Twins Environment for Energy Performance Optimisation, Self-consumption Maximisation and Predictive Maintenance (BEPO)

7.5.1 Brief Description

The aim of the BEPO toolbox is to provide analytics and services for optimizing the operation and energy management of buildings in real-time.

This building analytics and management toolbox will comprise a set of following features

- Monitoring in real-time the energy performance of buildings as a whole but also at individual systems level;
- Analysing energy performance of building against the needs and requirements of building occupants;
- Supporting of energy management decision-making for optimizing it;
- In presence of RES, storage and other flexibility sources, the toolset will allow for the design of appropriate flexibility control strategies in order to maximise self-consumption and, thus, reduce energy costs, without compromising key occupants' comfort and well-being;
- Finally, the toolset will be further enhanced with predictive maintenance features for large energy consuming equipment and systems installed in buildings.

The toolset will be based on the Digital-Twin model, which is created based on a mix of existing mathematical & physical model of the building which will be trained based on real-life measurement data (indoor temperature, energy consumption of single consumption points, building automation data, IoT sensors data and weather data) along with data referring to the geometry of the building, the thermal properties of building envelope elements and flexibility analytics delivered through the analysis performed in the Analytics Toolkit of BEYOND.

This toolset will complement the developments of T5.3, with focusing on real-operation, appropriately extending the existing Digital Twin Model with context-aware flexibility profiles of occupants in the built environment, resulted from WP4. Finally, regarding predictive maintenance features, toolset will leverage the baseline predictive maintenance analytics of BEYOND (WP4) and further enriching them to address low-level identification of various mechanical faults or discrepancies, anomalous behaviours and defective equipment/ materials, Indoor Air Quality (IAQ) violations, while recommending optimised maintenance actions, which will be based on the prediction of future equipment performance.

7.5.2 Functional and Non-Functional Requirements

The realisation of the functionalities of BEPO toolset comes with a list of functional requirements as listed in the table below:

TABLE 7-8 BEPO FUNCTIONAL REQUIREMENTS

Functional Requirement ID	Description	Relevant Use Case(s)
BEPO_001	Toolset should provide near-real time analytics to support monitoring the energy performance of facilities and buildings (kwh, kwh/m2)	UC_6.3
BEPO_002	Toolset should calculate different KPIs (such as energy generation, consumption, storage, HVAC components energy efficiency, air quality) based on the near-real time performance of building	UC_6.3
BEPO_003	Toolset should provide comparative visualisation of KPIs against EPC targets	UC_6.3
BEPO_004	Toolset should be able to visualise Energy performance data of all the assets, including real-time monitoring data, generation and storage (kW / per asset - trend graphs)	UC_6.3
BEPO_005	Toolset should forecast building energy performance (kW / per asset - trend graphs)	UC_6.3
BEPO_006	Toolset should provide energy management analytics and output different energy performance improvement management strategies (storage usage, REW usage, etc., taking into account e.g., energy costs, weather, and other parameters)	UC_6.3
BEPO_007	Toolset should access and visualise a list of HVAC systems available in the building	UC_6.4
BEPO_008	Toolset should analyse near-real time performance of HVAC systems and compare with configured HVAC system performance target indicators (energy efficiency)	UC_6.4
BEPO_009	Toolset should analyse the history data of HVAC performance (discrepancies, the causes of these discrepancies, and implemented corrective actions)	UC_6.4
BEPO_010	Toolset should provide predictive analytics and output the list of potential maintenance actions	UC_6.4
BEPO_011	Toolset should provide the visualisation of the assets of the building (incl. malfunctions) and available energy sources	UC_6.5
BEPO_012	Toolset should provide the self-consumption analytics and respective algorithms.	UC_6.5
BEPO_013	Toolset should be able to access near-real time data including energy consumption & demand forecasts	UC_6.5

BEPO_014	Toolset should be able to simulate alternative energy control strategies to support multi-objective decision analysis towards the optimization of self-consumption and to reduce the energy costs.	UC_6.5
BEPO_015	Toolset should provide building performance optimisation analytics based on control strategy simulation resulted from BEPO_014 simulations	UC_6.5
BEPO_016	Toolset should provide self-consumption optimisation analytics based on control strategy simulation resulted from BEPO_014 simulations	UC_6.5
BEPO_017	Toolset should be able to send optimal control strategies to the generation assets	UC_6.5
BEPO_018	Toolset should be able to send optimal control strategies to the energy consumption assets of the building	UC_6.5

In addition to the functional requirements presented above, a list of non-functional requirements has been elicited and presented in the following table:

TABLE 7-9 BEPO NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement ID	Description
BEPO_N_001	The tool shall be privacy aware (complaining with pseudonymised, anonymised as well as personal data principles)
BEPO_N_002	The tool shall be always up to date with the latest data available
BEPO_N_003	The tool shell should have login requirements
BEPO_N_004	The tool should be enterable with open standards
BEPO_N_005	The tool shall interact with BEYOND platform (data access & sharing)
BEPO_N_006	The tool shall provide an intuitive and user-friendly interface to the user
BEPO_N_007	The tool shell be documented
BEPO_N_008	The tool shall have sufficient performance in loading and update times
BEPO_N_009	The tool shall have high performance in the presentation of simulation results
BEPO_N_010	The tool should be compatible with design standards
BEPO_N_011	The tool shall be reliable, scalable and available

7.5.3 Architecture - Components View

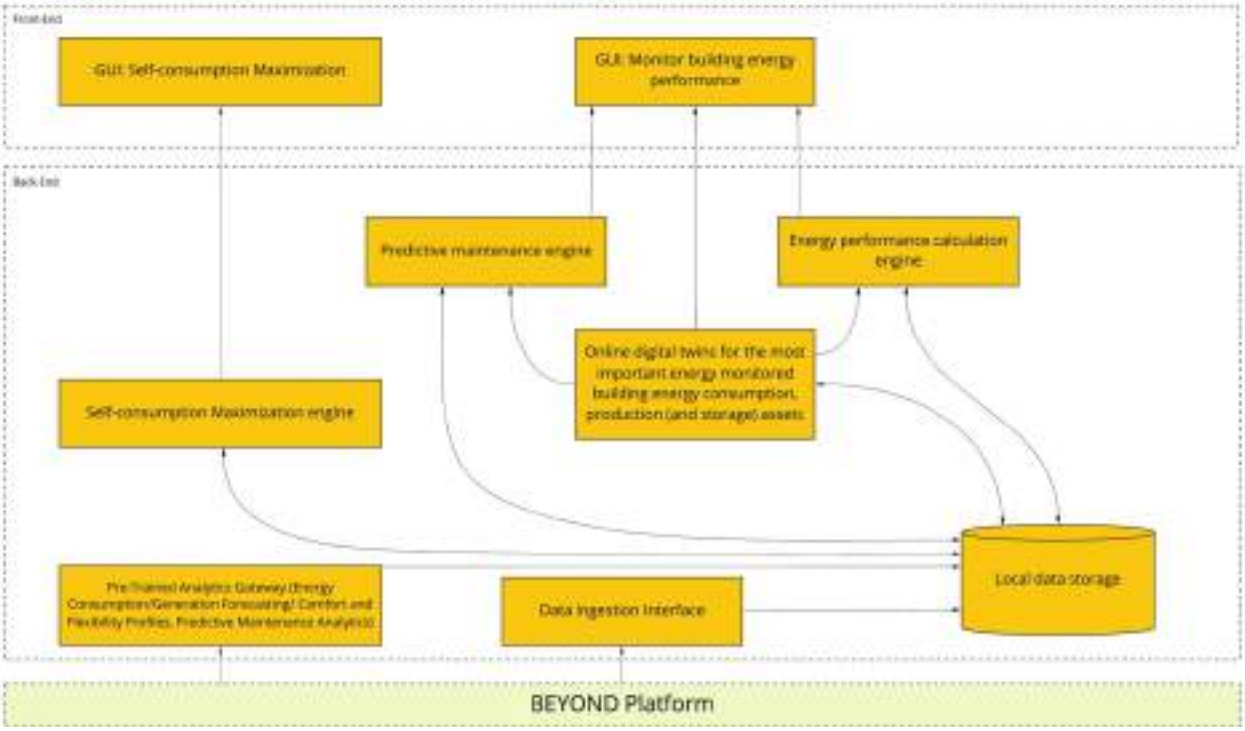

FIGURE 7-10: BEPO ARCHITECTURE

BEPO-1: GUI for monitoring building energy performance: This GUI makes it possible to display building energy performance and possible malfunctions and inefficiencies for supporting predictive maintenance.

BEPO-2: Digital twin for energy assets: These digital twins are machine learned on-line models that gives also same values as studied energy asset related energy measurement. Digital twins are possible to generate for the most important energy monitored building energy consumption, production and storage assets. These digital twins will be used for energy performance monitoring but can be utilised also e.g., for energy optimisation purposes.

BEPO-3: Predictive maintenance engine: This predictive maintenance support engine use machine learning technologies to find malfunctions and inefficiencies from selected building assets.

BEPO-4: Energy performance calculation engine: This software calculates building energy performance using measurement data and e.g., digital twin boosted International Performance Measurement and Verification (IPMVP) method. In addition, selected building energy assets' related KPIs (e.g., building space heating energy consumption as kWh/m^2, heat pump COP coefficient, heat recovery efficiency) will be calculated based on measurement data.

BEPO-5: Local data storage: This database includes in this case needed input and output values (optionally communicating with BEYOND platform in the background).

BEPO-6: GUI for self-consumption maximisation: This GUI support consumers in figuring out the best option for their unique situation towards self-optimizing energy management and understanding how their data and their flexibility can generate new income.

BEPO-7: Self-consumption maximisation engine: This software calculates basic information for end user how to maximise self-optimizing energy management.

7.4.5 Sequence Diagram

The flow diagram in below represents the interaction between the BEPO components, which were defined earlier. GUI components (for monitoring building energy performance and for self-consumption maximisation) are used to visualise results of the analytics to the end user.

From backend point of view, Digital Twin and Predictive Maintenance Engine components interacts with local data store to access necessary datasets and real-time measurements and store simulation results. Further the results of predictive maintenance analytics are communicated to GUI component to display the performance state of the building (incl. malfunctions, etc.)

Energy performance Engine and Self-consumption optimisation Engine components communicate with local data store to access required data sets in order to perform analytics. The results are stored back to the data store and also displayed to the end user on the GUI components (Energy management and Self-consumption Optimisation component).

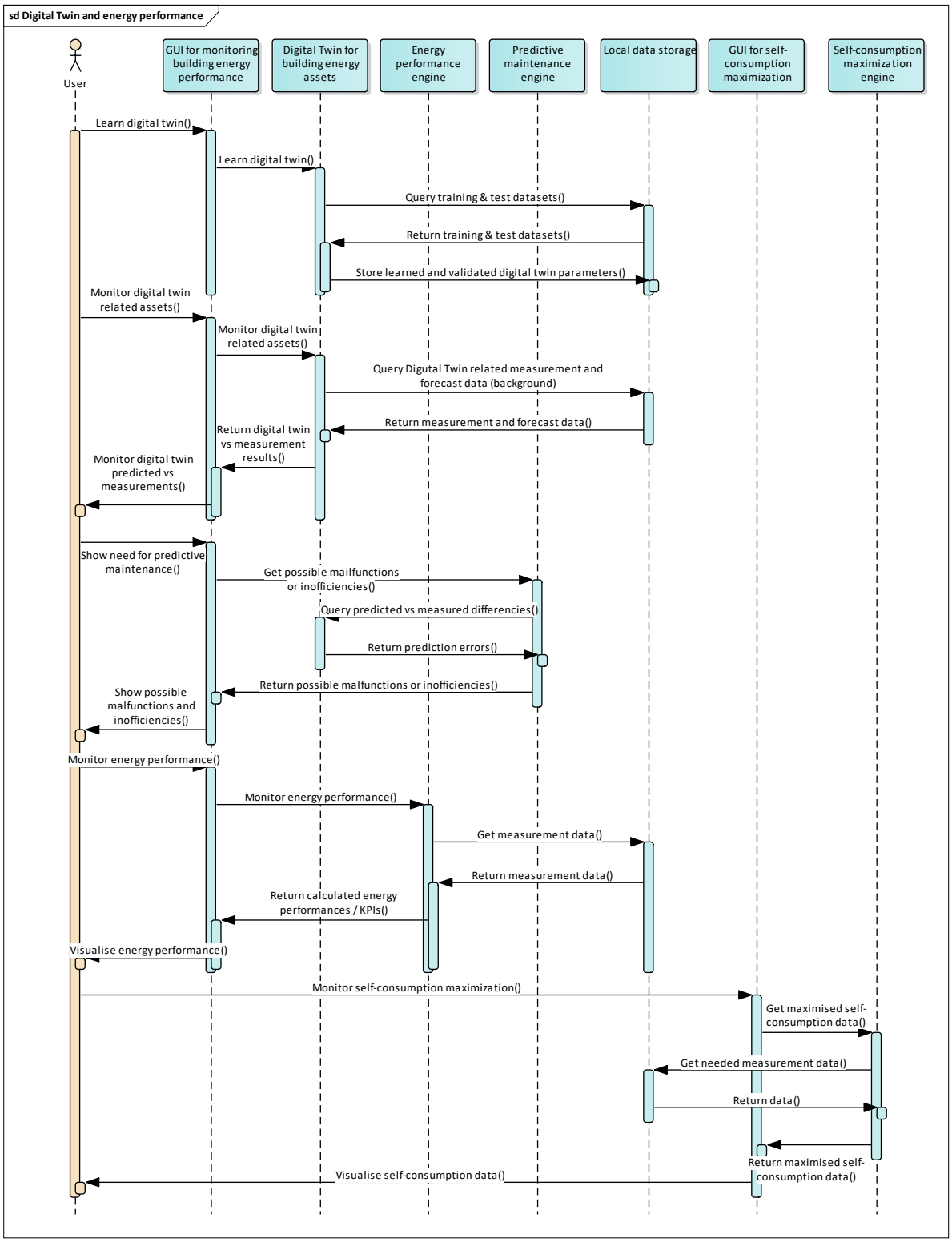

FIGURE 7-11: BEPO FLOW DIAGRAM

7.6 Building Portfolio Management Optimisation Tool (BPMO)

7.6.1 Brief Description

The Building Portfolio Management Optimisation Tool (BPMO) will offer a bundle of services to the retailers aiming at:

1. The development, presentation and utilisation of retailer portfolio analytics for elasticity estimation and extraction of useful insights.
2. The elasticity utilisation in implicit Demand Response and the development of imbalance reduction strategies.

The main scope of the Tool is to enable the sensible management of the retailers' portfolio according to the needs for improved monitoring and adjusting of the consumers' demand in order to minimise the costly imbalances of the retailers. The Tool will manage the portfolio seeking and categorizing according to spatio-temporal patters, segmenting and clustering in a manner that will enable the development of policies and strategies for grouped customers, under specified criteria.

The exploitation of the implicit flexibility through the utilisation of dynamic pricing can be achieved with improved accuracy as long as the pricing is better suited to the need of each customer category derived by elasticity to the price fluctuation. Thus, the price determination and the corresponding signal methods to the customer can both load shift for the reduction of the retailer imbalances, but also to increase the profitability of the retailer.

Moreover, the portfolio management provided by the Tool will also be utilised for means of Energy Efficiency improvement of the retailers' portfolio. The EU members are imposed to EU Energy Efficiency obligations which are passed through to the local retailers regarding the final consumption of their portfolios. The clustering of the portfolios can ensure the improved accuracy of the developed Energy Efficiency strategies, as well as the better monitoring of their results. In that sense, the clustering of the retailers' portfolios can achieve:

- An accurate and continuous monitoring of the portfolio Energy Efficiency regarding the goals and thresholds imposed by the EU Energy Efficiency obligations.
- The development of strategies to motivate the Energy Efficiency of their portfolio through better-valued investments and influencing their customers' energy behaviour, especially through decreasing their demand during peak hours.

More specifically, concerning the billing strategies, there are various pricing schemes[25] that have been proposed for Demand-Side Management (DSM), which are either

[25] Vardakas, J. et al. "A Survey on Demand Response Programs in Smart Grids: Pricing Methods and Optimisation Algorithms." IEEE Communications Surveys & Tutorials 17 (2015): 152-178.

retail price structures or DR-based programs. The BPMO tool deals with the former case where either fixed prices or consumption-based electricity rates are offered to consumers in order to motivate the decrease in their consumption especially in the peak time slots. It will offer one or more of the schemes described below:

Flat pricing has been and is still being used in traditional energy systems. Under this scheme, the only way to reduce the electricity bills is by simply using less electricity throughout the duration of the day. Although the flat pricing is generally non dynamic, in some cases seasonal flat pricing can be applied, where prices are fixed within a season, but they can change from one season to another[26].

Time-Of-Use (TOU) pricing is more dynamic since it is the application of flat pricing in different time periods. Under a TOU pricing scheme, prices are fixed within different pricing periods, which can be different hours within a day or different days within a week[27]. However, the effectiveness of such schemes to the reduction of the total power consumption is supposed to be limited, since customers do not receive any practical incentives to reduce or shift their demands. Their response to TOU schemes is triggered by the fact that they receive attractive off-peak prices, but relatively high prices in peak-demand hours. A study[28] showed that TOU programs offer the smallest reduction in the peak demand among all tested programs.

Critical Peak Pricing (CPP) is similar with TOU pricing, regarding the fixed prices in different time periods. However, the specific price for at least one period can change, either regularly or in most cases, due to occasions of system stress[29]. The participating consumers *receive notification of the new energy price, usually a day ahead*. As in the case of TOU, CPP is less economically efficient for the consumers, due to the preset prices. Obviously, the ratio of on-peak to off-peak price is higher on CPP event days than in a TOU program. The energy suppliers can use CPP to achieve significant load reductions during critical periods, but with high probability of negative net benefits. A variation of CPP is the **Extreme Day CPP**, in which a critical peak price is applied to critical peak hours, but there is no variable tariff on other days[30].

---

[26] M. Doostizadeh and H. Ghasemi, "A day-ahead electricity pricing model based on smart metering and demand-side management", Energy, vol. 46, no. 1, pp. 221-230, October 2012.
[27] J. Aghaei and M.-I. Alizadeh, "Demand response in smart electricity grids equipped with renewable energy sources: A review", Renew. Sustain. Energy. Rev., vol. 18, pp. 64-72, Feb. 2013.
[28] A. Faruqui, S. Sergici, A. Sharif, "The impact of informational feedback on energy consumption - A survey of the experimental evidence", Energy, vol. 35, no. 4, pp. 1598-1608, Apr. 2010.
[29] Q. Zhou, W. Guan, and W. Sun. "Impact of demand response contracts on load forecasting in a smart grid environment", in Proc. 2012 PES Gen. Meet., pp.1-4, San Diego, CA, USA, 22-26 Jul. 2012.
[30] A. Faruqui and S. S. George, "The value of Dynamic Pricing in Mass Markets", Electricity J., vol. 15, no. 6, pp. 45-55, Jul. 2002.

In **Peak Load Pricing (PLP)** each day has a number of time periods and different prices are determined for each period. These prices are announced to the customers *ahead of each day*[31]. The price calculation for each time period is based on the average power consumption of the consumers in each time period, scoping the maximisation of the payoff of the energy provider[32]. Also, the price calculation targets on the peak demand shifting, by expecting a reaction from the customers' side according to the high price.

A similar but more dynamic pricing scheme is **the adaptive pricing**, where prices are not announced to customers at the beginning of day but, instead, *at the beginning of each time period*[33]. The price calculation is performed in real time and is based on the power consumptions on previous time periods.

Under the **Peak Day Rebates (PDR)** pricing scheme, customers decide whether they respond to a critical event. Specifically, customers are under their standard tariff, but they have the opportunity to receive a rebate payment for any load reduction they can achieve below an estimated baseline load threshold[34]. The results of a pilot study in USA, showed that PDR is more effective compared to TOU, in terms of power reduction and consumer's satisfaction[35]. On the other hand, the same study showed that CPP is more beneficial that PDR. Furthermore, due to the fact that the baseline load threshold must be calculated for each customer and for every critical event, additional resources are needed.

Another pricing scheme is the **Vickrey-Clarke-Groves (VCG)** scheme. According to it, customers are requested to provide their power demand information, which is then utilised by a central mechanism for the price calculation, for each time period[36]. Payments are provided to the customers in a way that they have motivations to provide their demand information truthfully.

Real-Time Pricing (RTP) requires the maximum customer participation. Under an RTP scheme, the energy provider announces electricity prices on a rolling basis; these prices are determined and announced *before the start of each time period (e.g., 15*

[31] M. Crew, C. Fernando, and P. Kleindorfer, "The theory of peak-load pricing: A survey", J. Regulatory Econ., vol. 8, no. 3, pp. 215-248, Nov. 1995.
[32] P. Samadi, H. Mohsenian-Rad, R. Schober, and V. W. Wong, "Advanced demand side management for the future smart grid using mechanism design". IEEE Trans. Smart Grid, Vol 3, no. 3, pp. 1170-1180, Sept. 2012.
[33] P. Samadi, A. Mohsenian-Rad, R. Schober, V. Wong, and J. Jatskevich, "Optimal real-time pricing algorithm based on utility maximisation for smart grid", in Proc. IEEE SmartGridComm 2010, pp. 415-420, Gaithersburg, Maryland, USA, 4-6 Oct. 2010.
[34] M. Doostizadeh and H. Ghasemi, "A day-ahead electricity pricing model based on smart metering and demand-side management", Energy, vol. 46, no. 1, pp. 221-230, October 2012.
[35] J. Wang, M. A. Biviji, and W. Maria Wang, "Lessons learned from smart grid enabled pricing programs", in Proc. IEEE PECI 2011, Urbana, IL, USA, 25-26 Feb. 2011.
[36] P. Samadi, A. Mohsenian-Rad, R. Schober, V. Wong, and J. Jatskevich, "Optimal real-time pricing algorithm based on utility maximisation for smart grid", in Proc. IEEE SmartGridComm 2010, pp. 415-420, Gaithersburg, Maryland, USA, 4-6 Oct. 2010.

minutes beforehand)[37]. In that sense, the successful implementation of an RTP scheme relies on the two-way communication capabilities of the smart grid, which together with an Energy Management Controller (EMC) installed at the customer's premises, significantly increases the decision taking speed[38]. EMCs enable continuous data flow and are calibrated according to the consumer's preferences. The consumers make smart decisions to modify the energy usage across the building, which will guarantee higher reductions in the electricity bill. The energy provider also makes decisions to define the prices for the upcoming time period. These decisions are influenced by random events, the total power consumption and the response of the consumers to the previous prices[39]. Although RTP mechanisms have already been applied to large industrial and commercial customers, in the residential domain they have a smaller implementation success, since most consumers are risk-averse

and see the necessity of taking systematic electricity decisions as an important drawback[40,41]. In addition, in some cases the cost savings resulting from the participation in an RTP program will exceed the costs imposed on customers to follow the program[42].

7.6.2 Functional and Non-Functional Requirements

Below is a list of the functional requirements that BPMO tool will satisfy:

TABLE 7-10 BPMO FUNCTIONAL REQUIREMENTS

Functional Requirement ID	Description	Relevant Use Case(s)
BPMO_001	The Beyond Building Portfolio Management Optimisation Tool shall support the clustering of customers based on criteria specified by the retailer (level of demand, demographics, specific tariff, location, elasticity to price fluctuations)	UC_6.10
BPMO_002	The Beyond Building Portfolio Management	UC_6.10

[37] C. Chen, S. Kishore, and L. V. Snyder., "An innovative RTP-based residential power scheduling scheme for smart grids", in Proc. IEEE ICASSP, pp. 5956-5959, Prague, Czech Republic, 22-27 May 2011.
[38] G. Webber, J. Warrington, S. Mariethoz, and M. Morari, M. Communication limitations in iterative real time pricing for power systems, in Proc. IEEE SmartGridComm 2011, pp. 445-450, Brussels, Belgium, 17-20 Oct. 2011.
[39] P. Luh, Y. Ho, and R. Muralidharan, "Load adaptive pricing: An emerging tool for electric utilities,"IEEE Trans. Autom. Control, vol. AC-27, no. 2, pp. 320-329, Apr. 1982.
[40] S. M. Baladi, J.A. Herriges, and T. J. Sweeney, "Residential response to voluntary time-of-use electricity rates,"Resour. Energy Econ., vol. 20, no. 4, pp. 225-244, Sept. 1998.
[41] H. Allcott, "Real time pricing and electricity markets", Working Paper, Harvard Univ., Feb. 2009.
[42]] H. Allcott, "Real Time Pricing and Electricity Market Design", New York University Working Paper, May 2012.

	Optimisation Tool shall support the creation of optimal energy demand curves per individual/group/portfolio.	
BPMO_003	The Beyond Building Portfolio Management Optimisation Tool shall support the calculation of price elasticity per individual/group/portfolio.	UC_6.10, UC_6.11
BPMO_004	The Beyond Building Portfolio Management Optimisation Tool shall support the calculation of a priori and a posteriori profit for the retailer due to the implementation of dynamic pricing compared to BaU.	UC_6.10, UC_6.11
BPMO_005	The Beyond Building Portfolio Management Optimisation Tool shall support the comparison of the a priori and a posteriori overall cost for each cluster after the implementation of dynamic pricing compared to BaU.	UC_6.10
BPMO_006	The Beyond Building Portfolio Management Optimisation Tool shall provide forecasting of the day-ahead retail prices.	UC_6.10, UC_6.11
BPMO_007	The Beyond Building Portfolio Management Optimisation Tool shall provide signal dispatch to customers based on various pricing algorithms (e.g., Time of Use, Real Time Pricing, Critical Peak Pricing).	UC_6.10, UC_6.11
BPMO_008	The Beyond Building Portfolio Management Optimisation Tool shall provide visualisation of customer information at different levels, individual, group and portfolio.	UC_6.10, UC_6.11
BPMO_009	The Beyond Building Portfolio Management Optimisation Tool shall provide visualisation of energy consumption and production at building level (both historical and day-ahead forecasted).	UC_6.10
BPMO_010	The Beyond Building Portfolio Management Optimisation Tool shall provide visualisation of historical and day-ahead wholesale energy prices.	UC_6.10, UC_6.11
BPMO_011	The Beyond Building Portfolio Management Optimisation Tool shall provide visualisation of historical and day-ahead retail energy prices.	UC_6.10, UC_6.11
BPMO_012	The Beyond Building Portfolio Management Optimisation Tool shall provide the option to retailers to add the pricing schemes reductions for the customer participation in the implicit DR programs.	UC_6.10
BPMO_013	The Beyond Building Portfolio Management Optimisation Tool shall optimise the customer clusters based on information received by the Personalised Energy Analytics Tool.	UC_6.10
BPMO_014	The Beyond Building Portfolio Management Optimisation Tool shall aid the retailers to minimise errors and imbalances arising from customer behaviour around price fluctuations.	UC_6.11

BPMO_015	The Beyond Building Portfolio Management Optimisation Tool shall enable the monitoring and quantification of activities by the retailer according to the EU Energy Efficiency Obligations.	UC_6.10
BPMO_016	The Beyond Building Portfolio Management Optimisation Tool shall aid the retailers to optimise the day-ahead bidding of electricity.	UC_6.11

In addition to the functional requirements presented above, a list of non-functional requirements has been elicited and presented in the following table:

TABLE 7-11 BPMO NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement ID	Description
BPMO_N_001	The Tool shall provide a friendly and intuitive UI to the end-users.
BPMO_N_002	The Tool shall provide minimum latency to the end-users.
BPMO_N_003	The Tool should be design aid users to understand the different features provided.
BPMO_N_004	The Tool shall provide access to the end-users via web.
BPMO_N_005	The Tool shall provide application localisation.

7.6.3 Architecture - Components View

The BPMO Tool consists of the following components:

- **BPMO-1: The Portfolio Analytics Dashboard** enables the creation and manipulation of a dashboard presenting multiple indicators and analytics of the retailer's portfolio. It is fed with information concerning the generation and consumption of the portfolio and the results of the dynamic price forecasting engine. Moreover, it presents helpful graphs from the appropriate clustering of the customer portfolio that enable the retailer to develop accurate strategies for each category of customer to achieve monitorable and predictable increase in the Energy Efficiency of their portfolio. Such graphs could present the consumption of each cluster before, during and after the implementation of an Energy Efficiency strategy in order to enable its evaluation.
- **BPMO-2: The Dynamic Price Forecasting Engine** applies the Dynamic Pricing strategy to utilise the implicit flexibility of the retailer's portfolio. It gains feedback from the Price Elasticity Engine, the Optimal Energy Demand Engine and the consumption/generation forecasts in order to determine the most suitable price for each timeslot.
- **BPMO-3: The Price Elasticity Engine** is responsible for the calculation and definition of the demand elasticity to price fluctuations of each identified cluster of the retailer's portfolio that is imposed to the Dynamic Pricing Model.

- **BPMO-4: The Optimal Energy Demand Engine** calculates and determines the optimal energy demand regarding the wholesale price of the energy as purchased in the energy market by the energy retailer. The optimal demand curve is crucial for the determination of the energy price posed to the retailer's clusters of customers by the retailer.
- **BPMO-5: The Customer Clustering Engine** is responsible for the multiple clustering and segmentation of the retailer's portfolio according to various needs regarding pricing and energy efficiency strategies.
- **BPMO-6: The pre-trained Analytics Gateway (Energy Consumption/Generation Forecasting)** is fed with data and analytics from the BEYOND cloud-based platform regarding consumption and generation forecasting for the retailers' customers.
- **BPMO-7: The Data Ingestion Interface** is used to retrieve from the BEYOND Cloud-based platform data as well as ready-made analytics and models, that are locally stored and fed into the Dynamic Price Forecasting Engine.
- **BPMO-8: The Local Data Storage** is a semi-persistent storage facility which is used to temporary hold data that come from the BEYOND platform in order to load them in the other components, and to store application relevant configuration settings, as well as logs regarding the operations performed by the other components in a persistent manner.
- **BPMO-9: The Data Exposure Interface** locally stores data analytics that would be useful for the tasks and activities performed by the BPMO application, especially for the profiling of the customers.

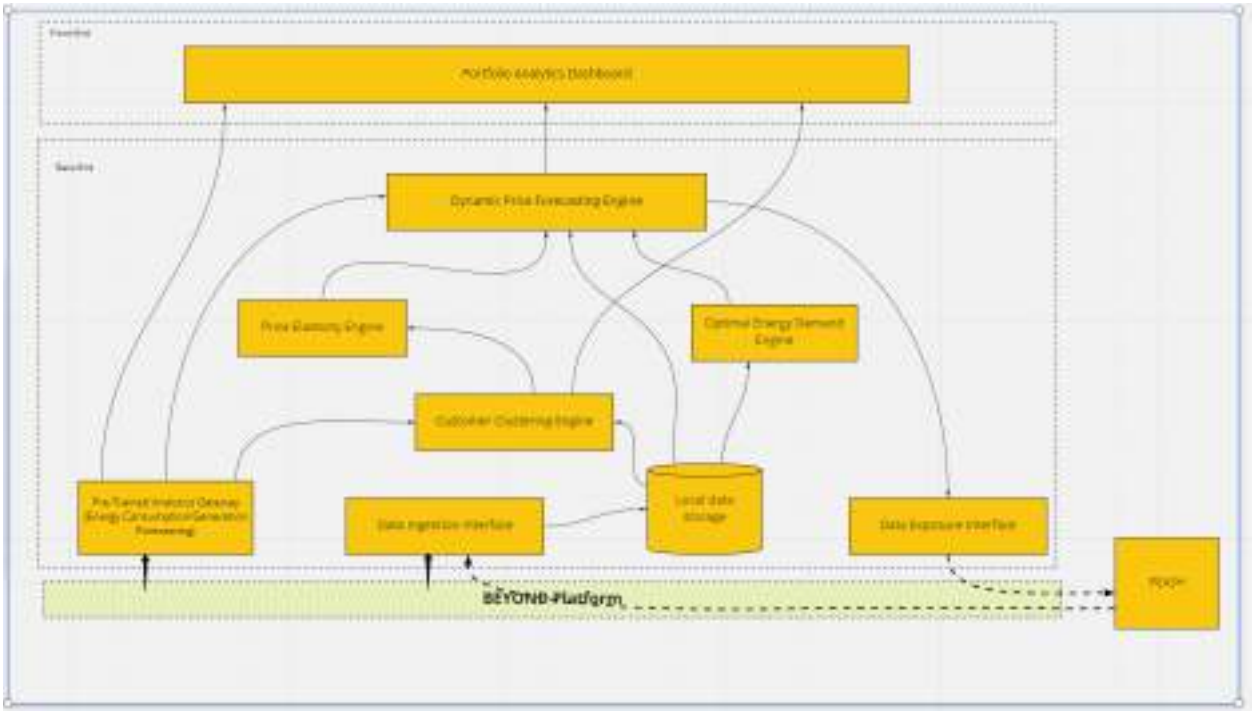

FIGURE 7-12: BPMO ARCHITECTURE

7.6.4 Sequence Diagram

The following figure presents the flow diagrams of how a user is operating with the BPMO application.

Scenario 1: As shown in the sequence diagram, a subscribed user (energy retailer) can visit the dashboard to gain information and analytics based on historic data of their portfolio in various levels (individual, cluster or portfolio). According to the required criteria, they can have access to information of clustered groups of their portfolio regarding their demand profiles, their elasticity to price fluctuations, etc.

Scenario 2: Moreover, the retailer can ask for forecasted information about their portfolio. The dashboard will, then, communicate with the pre-trained Analytics Gateway in order to provide the generation and consumption predictions.

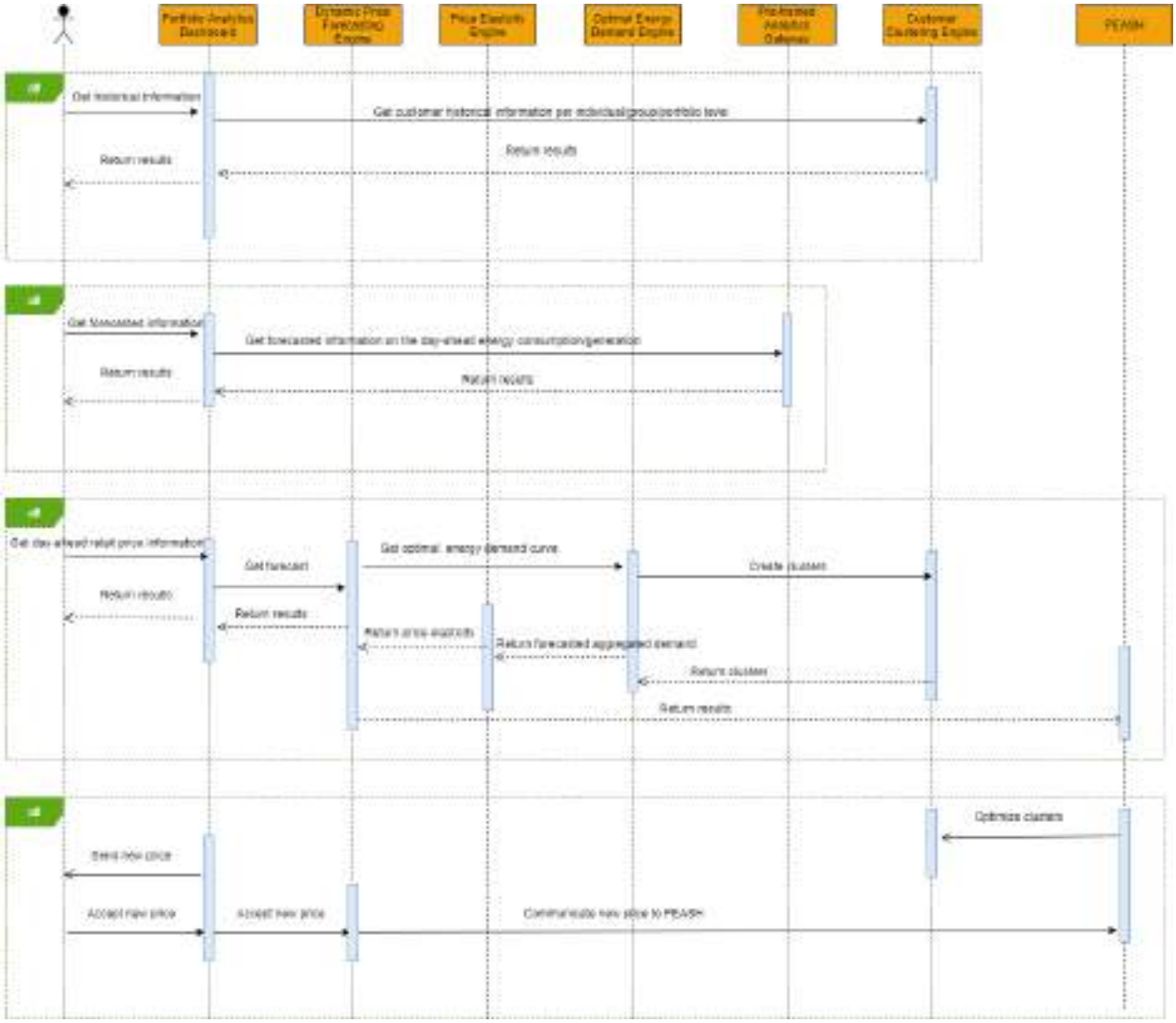

FIGURE 7-13: BPMO FLOW DIAGRAM

Scenario 3: The retailer can use the application to calculate and determine the retail prices based on the day-ahead market prices. Specifically, the wholesale prices will then be used for the determination of the optimal demand curve. The customers in the portfolio can be clustered according to their elasticity to energy price fluctuations. The prices for each cluster will be calculated according to their corresponding optimal demand curve. The output will become available not only to the Portfolio Analytics Dashboard but also to the PEASH application, through the Data Exposure Interface.

Scenario 4: In the feedback scenario, the response of the customers to the prices will affect the price elasticity of each customer or each cluster of customers. The new individual customer or cluster elasticity will be utilised for the future price determinations, so that steps from scenario 3 can be repeated.

7.7 Personalised Energy Analytics Tool for Guidance on Energy Performance Optimisation and Human-Centric Control Automation (PEASH)

7.7.1 Brief Description

The Personalised Energy Analytics for Guidance on Energy Performance Optimisation and Human- Centric Control Automation (PEASH) Application will offer a bundle of ambient and personalised services to the building occupants aiming at:

1. Increasing prosumer awareness about their energy behaviour patterns and their energy consumption within the built environment through personalised energy analytics,
2. Engaging prosumers in a continuous process for improving their energy behaviour through the utilisation of appropriate behavioural triggering techniques (behavioural segmentation and normative comparison).
3. Providing guidance in the form of tailored and personalised messages for triggering control actions on specific loads for optimizing energy performance and increasing energy savings.

The application will be fed with and continuously adapt to real-time (or near real-time) information captured by smart meters, sub-meters and sensors available in a building. Moreover, it will step on valuable knowledge offered from the BEYOND Data Analytics Containers referring to Comfort Profiles of Building Occupants and associated Context-Aware Flexibility Profiles, to render timely, context-aware, intuitive and personalised feedback through different modalities and channels thus ensuring that any guidance and control action is performed in a human-centric manner.

Moreover, the application will incorporate advanced functionality for intelligent automated control of building assets, enabling:

- The realisation of Energy Performance Optimisation Strategies through Scheduling the operation of building loads, or even more, remotely controlling them.
- The realisation of the derived strategies in a non-intrusive manner through smart automation always in respect of the comfort preferences and well-being needs of the building occupants.

The human-centric approach followed for the definition of Energy Performance Optimisation strategies will also enable the realisation of Non-energy Services for buildings, focusing on comfort preservation and well-being enhancement (through human-centric control optimisation of HVAC devices), while incorporating security services enabled through scheduling of lighting devices' operation during night hours or absence periods.

7.7.2 Functional and Non-Functional Requirements

The realisation of the functionalities of the Personalised Energy Analytics for Guidance on Energy Performance Optimisation and Human- Centric Control Automation Application comes with a list of functional requirements as listed in the table below:

TABLE 7-12 PEASH FUNCTIONAL REQUIREMENTS

Functional Requirement ID	Description	Relevant Use Case(s)
PEASH_001	The application shall provide a dashboard for the visualisation of energy demand data	UC_6.6
PEASH_002	The application shall provide a diagram for the visualisation of smart meter timeseries data	UC_6.6
PEASH_003	The application shall provide diagrams for the visualisation of individual devices energy consumption timeseries data	UC_6.6
PEASH_004	The application shall provide a histogram for the visualisation of daily energy consumption data (per hour)	UC_6.6
PEASH_005	The application shall provide a histogram for the visualisation of weekly energy consumption data (per day)	UC_6.6
PEASH_006	The application shall provide a histogram for the visualisation of monthly energy consumption data (per day)	UC_6.6
PEASH_007	The application shall provide a histogram for the visualisation of yearly energy consumption data (per month)	UC_6.6
PEASH_008	The application shall enable zooming in and out of histograms and timeseries data visualisations	UC_6.6
PEASH_009	The application shall enable the selection of the timeframe applied over the visualisations	UC_6.6
PEASH_010	All visualisations of the application shall be available in a dashboard	UC_6.6
PEASH_011	The application shall provide a dashboard for the visualisation of energy generation data	UC_6.6
PEASH_012	The application shall provide a histogram for the visualisation of daily energy generation data (per hour)	UC_6.6
PEASH_013	The application shall provide a histogram for the visualisation of weekly energy generation data (per day)	UC_6.6
PEASH_014	The application shall provide a histogram for the visualisation of monthly energy generation data (per day)	UC_6.6
PEASH_015	The application shall provide a histogram for the visualisation of yearly energy generation data (per	UC_6.6

	month)	
PEASH_016	The application shall provide a graphical view (timeseries) of the energy balance at building level	UC_6.6
PEASH_017	The application shall provide a graphical view of the self-consumption at building level at different time frames (day, week, month, year)	UC_6.6
PEASH_018	The application shall enable the comparative visualisation of energy demand of two different time periods in the same timeseries diagram	UC_6.6
PEASH_019	The application shall enable the comparative visualisation of energy generation of two different time periods in the same timeseries diagram	UC_6.6
PEASH_020	The application shall enable the comparative visualisation of daily energy demand (at building or device level) for two different days at the same histogram	UC_6.6
PEASH_021	The application shall enable the comparative visualisation of weekly energy demand (at building or device level) for two different weeks at the same histogram	UC_6.6
PEASH_022	The application shall enable the comparative visualisation of monthly energy demand (at building or device level) for two different months at the same histogram	UC_6.6
PEASH_023	The application shall enable the comparative visualisation of daily energy generation for two different days at the same histogram	UC_6.6
PEASH_024	The application shall enable the comparative visualisation of weekly energy generation for two different weeks at the same histogram	UC_6.6
PEASH_025	The application shall enable the comparative visualisation of monthly energy generation for two different months at the same histogram	UC_6.6
PEASH_026	The application shall enable clustering of energy prosumers based on demographic characteristics	UC_6.6
PEASH_027	The application shall enable clustering of energy prosumers based on building/ apartment size	UC_6.6
PEASH_028	The application shall provide peer comparisons (prosumers of similar demographics and building/ apartment size) of total energy demand for different time frames	UC_6.6
PEASH_029	Peer comparisons shall visualise the energy demand of the application user, in comparison with the average of similar users, the best performing one and the lowest performing one	UC_6.6
PEASH_030	The application shall provide peer comparisons (prosumers of similar demographics and building/	UC_6.6

	apartment size) of additional KPIs (generation, CO2 emissions, energy cost, self-consumption) for different time frames	
PEASH_031	Peer comparisons shall visualise additional KPIs referring to the application user, in comparison with the average of similar users, the best performing one and the lowest performing one	UC_6.6
PEASH_032	The application shall provide a timeseries visualisation of energy demand forecast for the next 72 hours	UC_6.6
PEASH_033	The application shall provide a timeseries visualisation of energy generation forecast for the next 72 hours	UC_6.6
PEASH_034	The application shall provide a timeseries visualisation of energy self-consumption forecast for the next 72 hours	UC_6.6
PEASH_035	The application shall utilise context-aware demand flexibility profiles to provide tailored recommendations for energy savings	UC_6.6 UC_6.7
PEASH_036	Recommendations for energy savings should be provided in the form of specific control actions over monitored loads	UC_6.6 UC_6.7
PEASH_037	Recommendations for energy savings shall respect the comfort preferences of the user	UC_6.6 UC_6.7 UC_6.12
PEASH_038	The application should detect energy demand outliers	UC_6.6
PEASH_039	The application shall utilise appropriate means to visualise detected outliers	UC_6.6
PEASH_040	The application shall provide notifications to the user once an outlier is detected	UC_6.6 UC_6.7
PEASH_041	The application shall enable the visualisation and monitoring of the status of building devices	UC_6.6
PEASH_042	The application shall enable the remote alteration of the set-point of HVAC and lighting devices through a user-friendly interface	UC_6.6 UC_6.7
PEASH_043	The application shall allow the user to opt-in automated control features	UC_6.7 UC_6.12
PEASH_044	The application shall allow the user to schedule the operation and setpoints of specific loads through a calendar feature	UC_6.6 UC_6.7
PEASH_045	The application shall perform automated control actions over specific loads, considering the comfort profiles and context-aware demand flexibility profiles of users	UC_6.7 UC_6.12
PEASH_046	The application shall allow the user to select the devices that will be involved in automated control strategies	UC_6.6 UC_6.7 UC_6.12
PEASH_047	The application shall allow the user to select between alternative automation modes	UC_6.6 UC_6.7

		UC_6.12
PEASH_048	The application shall offer different automated control modes involving maximum energy efficiency, balanced energy efficiency/ comfort, maximum comfort	UC_6.7 UC_6.12
PEASH_049	The application shall provide notifications once a control action has been triggered	UC_6.6 UC_6.7 UC_6.12
PEASH_050	The application shall keep a log of all control actions performed, distinguishing between manual, scheduled and automated	UC_6.7 UC_6.12
PEASH_051	The application shall provide pre-defined automated control strategies for security	UC_6.7 UC_6.12
PEASH_052	The application shall allow the user to alter the pre-defined automated control strategies for security	UC_6.6 UC_6.7 UC_6.12
PEASH_053	The application shall continuously monitor and assess the comfort of users (under the selected strategy) and perform corrective automated controls over their loads to satisfy their functional objective (based on input coming from PEASH_048)	UC_6.7 UC_6.12

In addition to the functional requirements presented above, a list of non-functional requirements has been elicited and presented in the following table:

TABLE 7-13 PEASH NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement ID	Description
PEASH_N_001	The application shall have high availability
PEASH_N_002	The application shall be always up to date with the latest data available
PEASH_N_003	The application shall allow for user registration and login with credentials
PEASH_N_004	The application shall interconnect with the BEYOND Core Platform
PEASH_N_005	The application shall retrieve data assets (raw data and analytics) from the BEYOND Core Platform
PEASH_N_006	The application shall provide an intuitive and user-friendly interface to the user
PEASH_N_007	The application shall be accessible through desktop and mobile devices
PEASH_N_008	The application shall offer easy navigation through the different screens and functions
PEASH_N_009	The application shall have high performance
PEASH_N_010	The application shall provide appropriate logging mechanisms for all user actions

PEASH_N_011	The application shall provide the mechanisms for system upgrade with minimum downtime

7.7.3 Architecture - Components View

The PEASH Application consists in the following components:

- **PEASH-1: The Personalised Energy Analytics Engine,** enabling the analysis of energy data streams from buildings and assets available in buildings to extract meaningful insights for the end-user and trigger energy performance optimisation actions. This Engine will be a lightweight analytics engine, as most of the analytics tasks will be performed on the BEYOND Cloud based platform, and the results will be served to PEASH. This engine is the main component that will serve information to the frontend of the application (PEASH-2).
- **PEASH-2: The Energy Analytics Dashboard,** which enables the creation of customised dashboards for visualizing the results of the energy analytics in an intuitive and user-friendly manner.
- **PEASH-3: The Consumer Clustering Engine,** enabling the grouping and segmentation of buildings and consumers in multiple sub-groups based on common demographic, building and/ or energy behaviour parameters and characteristics. The output of this engine is provided to the Normative Peer Comparison Engine (PEASH-4).
- **PEASH-4: The Normative Peer Comparison Engine,** allowing the application of behaviour triggering techniques and methods based on peer pressure and comparison for energy behaviour alteration. The output of this engine is provided to the Personalised Energy Analytics Engine (PEASH-1).
- **PEASH-5: The Energy Savings Recommender,** which will be responsible for the definition of energy efficiency strategies and the generation of personalised recommendations for energy behaviour improvement through manual actions over specific building assets and loads, based on the processing of knowledge extracted from building data in terms of comfort profiles and context-aware demand flexibility profiles.
- **PEASH-6: The Notification Engine,** enabling the on-the-fly configuration of tailored and personalised notifications either for guiding consumers towards improving energy performance or for notifying them once a control action has been triggered automatically.
- **PEASH-7: The Building Asset Scheduler**, facilitating the configuration of assets' operation schedules based on suggestions provided by the Energy Savings Recommender, while giving the opportunity to the user to set up his/ her own asset control schedule.
- **PEASH-8: The Smart Building Automation Engine,** enabling the automated control of HVAC and lighting devices (and even more smart appliances if

available) in a non-intrusive manner, stepping on energy efficiency strategies generated by the Energy Savings Recommender.

- **PEASH-9: The Asset Remote Control Panel,** allowing the user to have a visual overview of the status and performance of their individual assets and remotely control them by defining the set-points that shall apply over them.
- **PEASH-10: The Data Ingestion interface**, which is used to retrieve from the BEYOND Cloud based platform data as well as ready-made analytics and models, that are fed into the PEASH Personalised Energy Analytics Engine.
- **PEASH-11: The Local Polyglot Storage**, a semi-persistent storage facility which is used to temporary hold data that come from the BEYOND platform in order to load them in the other components, and to store application relevant configuration settings, as well as logs regarding the operations performed by the other components in a persistent manner.

The architecture diagram of PEASH is shown in the next figure.

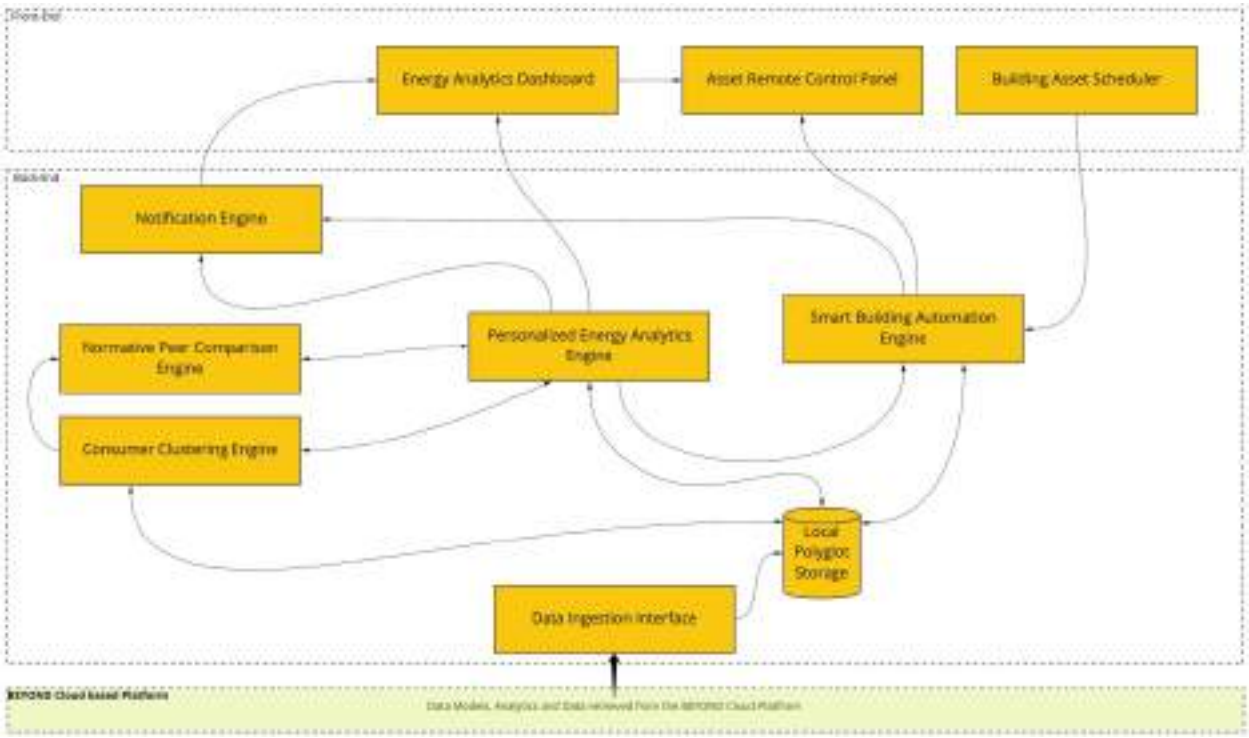

FIGURE 7-14: PEASH ARCHITECTURE

7.7.4 Sequence Diagram

The following figure presents a flow diagram of how a user is operating with the PEASH application.

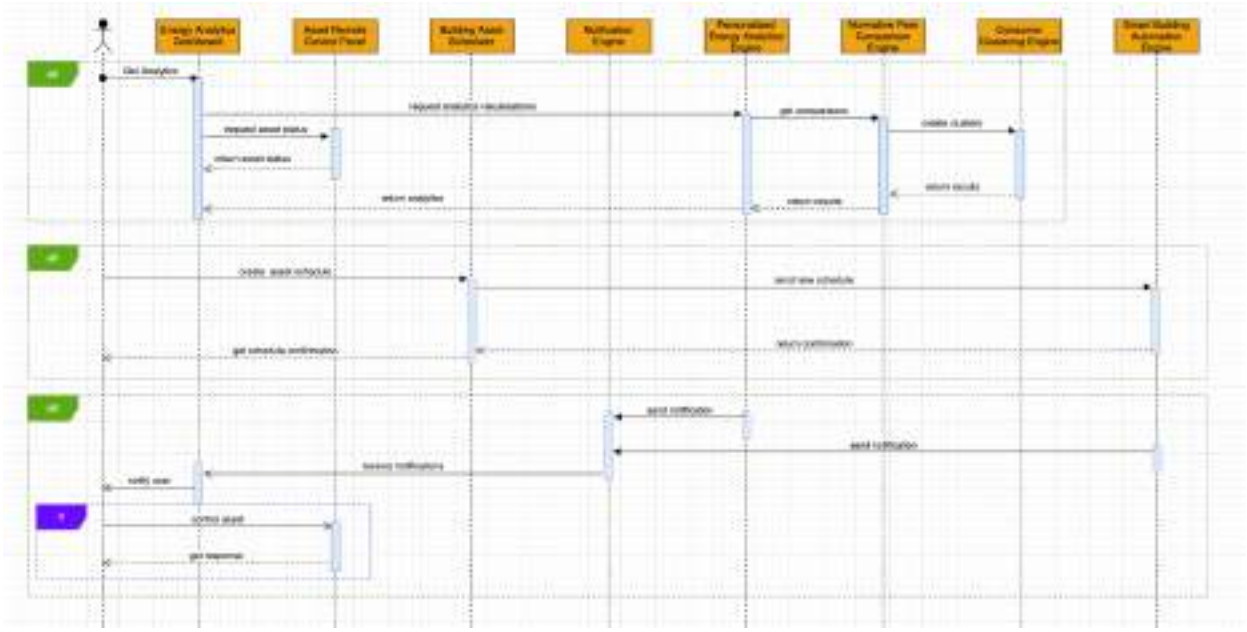

FIGURE 7-15: PEASH SEQUENCE DIAGRAM

As shown in the sequence diagram, a user can visit the dashboard and then he is able to get all the information regarding his case that is coming out of the Personalised Energy Analytics Engine as well as the status of his assets as this is coming out of the Asser Remote Control panel. Regarding the analytics, these are developed by getting the data stored in the BEYOND Cloud based Platform (not part of the sequence diagram), and then in turn the Personalised Energy Analytics Engine is returning the predefined analytics, while for the comparisons it is triggering with the Normative Peer Comparison Engine, which is getting the clusters out of the Consumer Clustering Engine in order to display the information.

Moreover, the use can also create a schedule for his/her assets. For that, he is engaging with the Building Asset Scheduler, which is providing the schedule then to the Smart Building Automation Engine, and a confirmation about the new schedule is sent back to the user.

Finally, in terms of Notifications, these are handled by the Notification Engine, which is receiving the necessary data in a push manner from the Personalised Energy Analytics Engine and from the Smart Building Automation Engine. Once notifications are provided to the user, the user can view them, and also take action regarding the operations of his/her assets, by engaging with the Asset Remote Control Panel

7.8 Energy Performance and Smart Readiness Certification Tool (EPSRC)

7.8.1 Brief Description

Traditional methods of building energy performance certification are mostly based on static data that does not reflect the dynamic elements of building operation and are not updated through time. The penetration of Energy Performance Contracts and the introduction of novel services for the building sector, require a more dynamic framework that will allow educated and evidence-based decision making on investments to be made, along with different investment incentives for different levels of energy performance certification levels over the building stock, through the utilisation of more granular data coming from the buildings.

Tailored tools and services are required towards the building sector allowing for their real time energy certification and investment optimisation through the utilisation of enhanced Display Energy Certificates that will allow for the buildings' compliance with energy efficiency commitments. Such services shall rely on real-life data streams and appropriate analytics that will be utilised for real-time monitoring of the energy performance of facilities/buildings (as a whole but also at individual systems level), analyse it against the needs and requirements of the occupants and visitors, define spatio-temporal outliers that significantly affect energy performance and support energy management decision-making for optimizing it.

Buildings are continuously gaining importance when it comes to the optimisation of the whole energy system and the provision of flexibility for improving the quality of power and participating in services offered to network operators. Though, in order to grasp this opportunity and be valid for participating in flexibility transactions, buildings need to achieve some minimum levels of smartness that can be certified through the use of the Smart Readiness Indicator Methodology.

In this sense, business actors like aggregators and retailers need to have access to detailed information about buildings towards evaluating their smart readiness and assessing their capabilities for getting involved in flexibility transactions and control strategies that are associated with innovative automation concepts. To this end, they need appropriate tools and methods to support the continuous assessment and certification of the Smart Readiness status of buildings. Such detailed information about building systems and other devices available, will need to be processed and analysed in order to enable the calculation of a variety of metrics that altogether contribute to the assessment of the smart readiness of the building and the presentation of individual analyses in appropriate graphical user interfaces (GUIs) that point out to the overall score, but also to the performance of the building in separate metrics that affect its smart readiness levels. The latter will point out to specific improvements that need to be performed (in terms of retrofitting) towards enhancing the smart readiness level of the building and making it more attractive for revenue

creation through flexibility provision (or at least for the provision of innovative smart energy services for energy savings).

7.8.2 Functional and Non-Functional Requirements

TABLE 7-14 EPSRC FUNCTIONAL REQUIREMENTS

Functional Requirement ID	Description	Relevant Use Case(s)
EPSRC_001	The system shall support capture of real-life data streams and appropriate analytics for monitoring in real-time the energy performance	UC_6.1
EPSRC_002	Display Energy Certificates shall consist in a rich dashboard displaying the ranking of the building/space as the key information and allowing for drilling into specific details analysis of energy consumption and performance outliers as defined by the following requirements.	UC_6.1
EPSRC _003	The system shall establish ranking of buildings/spaces to different operational ranking scales (A, B, C, D, E, F, G) in an automated manner by incorporation relevant baseline benchmarks provided by the industry.	UC_6.1
EPSRC_004	The system shall enable the visualisation of real-life data streams and analyse it against need and requirements,	UC_6.1
EPSRC_005	The system shall provide, analyse, and visualise a set of outliers that affect energy performance and support energy management decision making process	UC_6.1
EPSRC_006	The system shall gather general information on the building and its technical building systems	UC_6.2
EPSRC_007	The system shall identify the functionality level for each service running in the building	UC_6.2
EPSRC_008	The system should have visual presentation on calculation about Operational rating.	UC_6.2
EPSRC_009	The system shall provide possibility for user defined weighting factors	UC_6.2
EPSRC_010	The system shall make a list with all smart ready services and their functionality levels and provide possibility for users to include or exclude services	UC_6.2
EPSRC_011	The system shall provide with further details on the various services including the provisional impact scores for each of the services levels	UC_6.2
EPSRC_012	System shall provide different types of SRI	UC_6.2

	assessment defined by the user.	
EPSRC_013	System shall provide forms to the user to select the level of smartness of specific devices	UC_6.2
EPSRC_014	System should make sure that relevant methodology will be used in SRI calculation process	UC_6.2
EPSRC_015	System shall provide the ranking of the building according to the SRI calculation and achievement against specified benchmarks and thresholds	UC_6.2
EPSRC_016	System shall allow display of the different SRI assessments to allow comparison over the years.	UC_6.2
EPSRC_017	System shall offer comparative analysis of SRI values to similar consumers types, location, building size, building age etc.	UC_6.2

In addition to the functional requirements presented above, a list of non-functional requirements has been elicited and presented in the following table:

TABLE 7-15 EPSRC NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement ID	Description
EPSRC_N_001	The tool shall have high availability
EPSRC_N_002	The tool shall be always up to date with the latest data available.
EPSRC_N_003	The tool shall allow for user registration and login with credentials.
EPSRC_N_004	The tool shall interconnect with the BEYOND Core Platform.
EPSRC_N_005	The tool shall retrieve data assets (raw data and analytics) from the BEYOND Core Platform.
EPSRC_N_006	The tool interface shall be intuitive and user-friendly.

7.8.3 Architecture - Components View

The EPSRC Application consists of the following components:

- **EPSRC: Energy Assessment & Certification Tool** – allows the user to submit data and/or configuration parameters (weighting factors, types of assessment etc.) required for SRI calculation
- **EPSRC: Energy Performance Analytics Dashboard** – dashboard with various types of visualizations which display the calculated SRI score, analysed outliers and other relevant visualizations
- **EPSRC: Data Visualization Engine** – receives and processes requests from frontend applications, connects to local data storage to fetch required data and returns processed data to dashboards

- **EPSRC: Score Calculation Engine** - implements SRI calculation algorithms and stores the calculated results in local data storage
- **EPSRC: Data Analysis Engine** - locally stores data analytics that would be useful for the tasks and activities performed by the EPSRC application
- **EPSRC: User Configurations Engine** – processes and stores user defined settings, such as custom weighting factors, different types of SRI assessments, included/excluded services etc.
- **EPSRC: Data Ingestion Interface** - retrieves data from the BEYOND Cloud based platform as well as ready-made analytics and models
- **EPSRC: The Local Data Storage** is a semi-persistent storage facility which is used to temporarily hold the data that comes from the BEYOND platform for the purpose of loading it to other components, and to store application relevant configuration settings, as well as the logs regarding the operations performed by other components in a persistent manner.

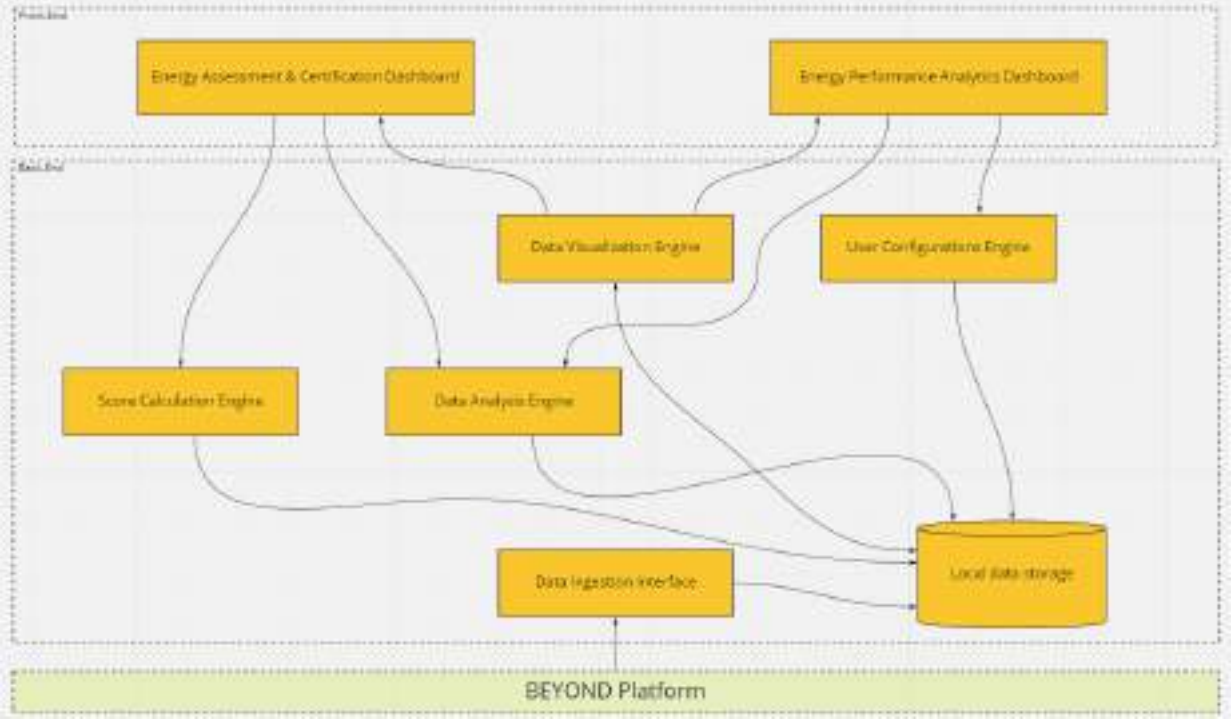

FIGURE 7-16: EPSRC ARCHITECTURE

7.8.4 Sequence Diagram

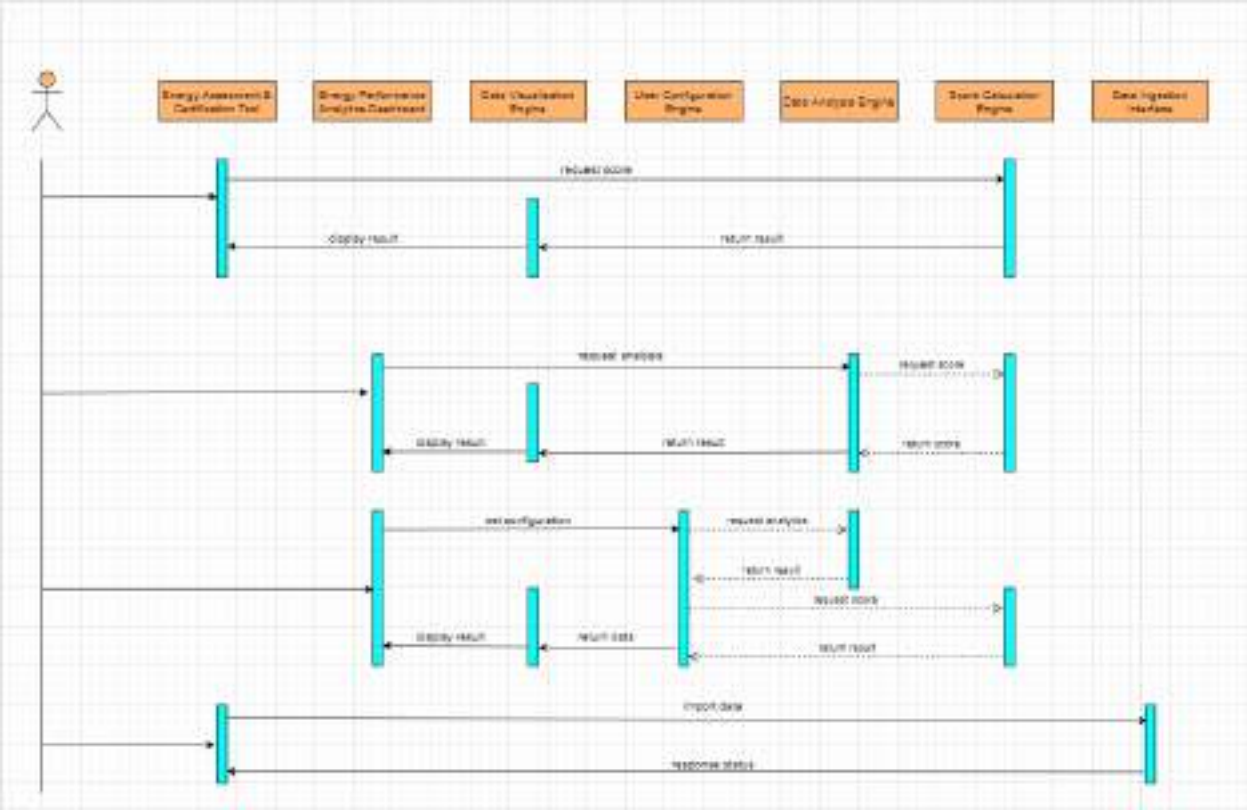

FIGURE 7-17: EPSRC FLOW DIAGRAM

As shown in the sequence diagram on **Error! Reference source not found.** a user can submit data and optionally configure parameters for SRI calculation. After data submission and SRI calculation, calculated score is returned to the dashboard and displayed to the user, along with other requested data visualizations. User can also review general building information data, request various data analysis (data outliers, service functionality levels, building rankings, comparisons between similar buildings etc.) and/or set additional configuration parameters (define weighting factors, include/exclude services, level of device smartness etc.).

7.9 Flexibility-based VPP Configurator and DR Strategy Optimisation Tool (FLEXopt)

7.9.1 Brief Description

The optimisation of flexibility delivered by building assets allows the creation of new direct revenue for building end-users, through provisioning the flexibility-based ancillary service to other market participants via aggregator intermediaries. The FLEXopt tool shall allow building managers to gain and maintain better insights into their flexibility portfolio, ensuring the utilisation of a maximum amount of flexibility that does not jeopardise the user comfort, nor does it threaten the proper functioning of the building assets.

Through the tool, the flexibility properly identified and selected by building managers can be delivered by means of an aggregator to third parties - flexibility users (i.e., system operators). The flexibility service user may be either the network operator seeking to resolve and optimise the operational network conditions or a balance group operator or another market participant.

On top of the delivered flexibility analytics, the tool shall perform flexibility optimisation analytics by providing flexibility segmentation, classification, and clustering of available flexibility resources. At the same time, the tool shall ensure a smooth operation of building assets under evolving conditions, enabling the optimal utilisation of flexibility and creation of revenue at the same time. The flexibility optimisation delivered by a portfolio of different building assets accompanied by Virtual Power Plant (VPP) configuration maximises the revenue creation space of the demand response aggregators, respecting at the same time operational and contractual limits.

The optimisation algorithms and VPP configuration models shall be directed towards meeting the request of the flexibility users, such as network operators, and shall consider: (i) flexibility characteristics (capacity, duration, ramp up/down time, shifting, shedding etc.) in terms of suitability for specific services; (ii) contractual parameters (from contracts signed between consumers/prosumers and aggregators) referring to number of activations, remuneration amount, etc. (iii) selected flexibility made available by consumers/prosumers.

7.9.2 Functional and Non-Functional Requirements

The functional requirements for the Flexibility-based VPP Configurator and DR Strategy Optimisation (FLEXopt) tool are listed below.

TABLE 7-16 FLEXOPT FUNCTIONAL REQUIREMENTS

Functional Requirement ID	Description	Relevant Use Case(s)
FLEXOPT_001	The tool shall display flexibility assets technical limitations towards delivering flexibility.	UC_6.8

FLEXOPT_002	The tool shall display base technical constraints of the assets with additional operational constraints derived from the impact on user comfort.	UC_6.8
FLEXOPT_003	The tool shall display consumption profiles per flexibility assets.	UC_6.8
FLEXOPT_004	The tool shall allow the selection of multiple flexibility assets.	UC_6.8
FLEXOPT_005	The tool shall display flexibility analytics per flexibility asset, taking into consideration its technical operational constraint and comfort-derived constraint imposed on this particular asset.	UC_6.8
FLEXOPT_006	The tool shall display flexibility analytics aggregated per selected flexibility assets taking into consideration technical and comfort constraints imposed on these assets.	UC_6.8
FLEXOPT_007	The tool shall display consumption profiles of aggregated flexibility assets.	UC_6.8
FLEXOPT_008	The tool shall allow the identification of flexibility requests by third parties e.g., network operators.	UC_6.8
FLEXOPT_009	The tool shall provide insights / visualisation of daily flexibility schedules delivered by aggregated flexibility assets.	UC_6.8
FLEXOPT_010	The tool may allow benchmarking of aggregated flexibility assets daily loads vs. their historical (baseline) daily consumption profiles.	UC_6.8
FLEXOPT_011	The tool may allow customisation of benchmark key performance indicators tailored to the users' needs.	UC_6.8
FLEXOPT_012	The tool may allow benchmarking of aggregated flexibility assets daily loads vs. their historical daily flexibility profiles.	UC_6.8
FLEXOPT_013	The tool shall allow flexibility segmentation of available flexibility sources delivered by building managers to aggregators.	UC_6.9
FLEXOPT_014	The tool shall allow flexibility classification of available flexibility sources delivered by building managers to aggregators.	UC_6.9
FLEXOPT_015	The tool shall provide flexibility clustering of available flexibility sources delivered by building managers to aggregators.	UC_6.9
FLEXOPT_016	The tool shall deliver optimised flexibility profiles to aggregators.	UC_6.9
FLEXOPT_017	The tool may provide customisation of criteria for flexibility optimisation to tailor the optimal flexibility profile to the users' needs.	UC_6.9
FLEXOPT_018	The tool shall allow the selection of flexibility sources delivered by building managers to aggregators.	UC_6.9

FLEXOPT_019	The tool shall provide optimal clustering of flexibility sources to aggregators.	UC_6.9
FLEXOPT_020	The tool shall display day-ahead electricity market or/and flexibility market requests to aggregators.	UC_6.9
FLEXOPT_021	The tool may provide real-time flexibility activation requests from third parties (i.e., DSOs, capacity market) to aggregators.	UC_6.9
FLEXOPT_022	The tool shall provide optimal VPP configuration of flexibility assets towards day-ahead market requests to aggregators.	UC_6.9
FLEXOPT_023	The tool may provide optimal VPP configuration of flexibility assets towards intra-day flexibility activation requests from third parties (i.e., DSOs, balancing capacity market) to aggregators.	UC_6.9
FLEXOPT_024	The tool shall provide continuous monitoring of the VPPs and re-configuration if overrides in flexibility signals occur (loss of flexibility source during a flexibility event)	UC_6.9
FLEXOPT_025	The tool may display intraday monitoring of energy or/and flexibility market requests to aggregators.	UC_6.9
FLEXOPT_026	The tool may allow benchmarking of aggregated flexibility assets daily loads vs. their historical daily flexibility profiles offered to third parties.	UC_6.9
FLEXOPT_027	The tool may allow customisation of benchmark key performance indicators of flexibility asset daily loads tailored to the users' needs (e.g., maximum flexibility activated, maximum savings, average savings etc.).	UC_6.9
FLEXOPT_028	The tool may provide insights of energy prices of day-ahead electricity, flexibility prices on the flexibility market or contracted transaction prices with third parties (i.e., DSOs, balancing capacity agreement with TSOs)	UC_6.9

In addition to the functional requirements presented above, a list of non-functional requirements has been elicited and presented in the following table:

TABLE 7-17 FLEXOPT NON-FUNCTIONAL REQUIREMENTS

Non-Functional Requirement ID	Description
NFLEXOPT_N_001	The tool shall have high availability
NFLEXOPT_N_002	The tool shall be always up to date with the latest data available.
NFLEXOPT_N_003	The tool shall allow for user registration and login with credentials.
NFLEXOPT_N_004	The tool shall interconnect with the BEYOND Core Platform.
NFLEXOPT_N_005	The tool shall retrieve data assets (raw data and analytics) from the BEYOND Core Platform.
NFLEXOPT_N_006	The tool interface shall be intuitive and user-friendly.

7.9.3 Architecture - Components View

The Flexibility-based VPP Configurator and DR Strategy Optimisation Tool consists in the following components:

- **FLEXopt: The Data Ingestion interface:** which is used to retrieve from the BEYOND Cloud based platform data as well as ready-made analytics and models.
- **FLEXopt: Pre-trained Analytics Gateway:** used to retrieve from the BEYOND Cloud based platform data ready-made analytics and models.
- **FLEXopt: Building manager Data Storage:** a semi-persistent storage facility which is used to temporary hold data that come from the BEYOND platform in order to load them in the other components, and to store relevant configuration settings, relevant for building managers, as well as logs regarding the operations performed by the other components in a persistent manner.
- **FLEXopt: Customer Portfolio Optimisation Engine:** enabling optimisation of technical and comfort constrains of building flexibility assets towards maximising the flexibility delivered to aggregators.
- **FLEXopt: Portfolio Analytics Dashboard for Building Managers:** enables the creation of customised dashboards for visualizing the results of the flexibility analytics in an intuitive and user-friendly manner for building managers
- **FLEXopt: Automatised Scheduler Engine:** enables automatised scheduling of flexibility assets within the building if the building manager activates such scheduler.
- **FLEXopt: Building Manager Assets Flexibility Scheduler:** permits the scheduling and selection of flexibility delivered by building assets from building managers to aggregators.
- **FLEXopt: Building Manager Assets Selection Engine:** enables the delivery of flexibility selected through the Scheduler from Building Managers to Aggregators.
- **FLEXopt: Aggregator Data Storage:** a semi-persistent storage facility which is used to temporary hold data that come from the BEYOND platform in order to load them in the other components, and to store tool relevant configuration settings, relevant for aggregators, as well as logs regarding the operations performed by the other components in a persistent manner.
- **FLEXopt: Aggregator Portfolio Optimisation Engine:** enables segmentation, classification and clustering of available flexibility sources delivered by building managers to aggregators and optimised flexibility profiles to aggregators.
- **FLEXopt: VPP Configuration Engine:** enables the VPP configuration of available flexibility towards intra-day flexibility activation requests from third parties (i.e.,

DSOs, balancing capacity market) to aggregators. Additionally, the tool shall provide continuous monitoring of the VPPs and re-configuration if overrides occur in flexibility signals.

- **FLEXopt: Portfolio Analytics Dashboard for Aggregators:** enables the creation of customised dashboards for visualizing the results of the flexibility analytics in an intuitive and user-friendly manner for aggregators.
- **FLEXopt: Aggregator Flexibility Scheduler:** permits the scheduling and selection of flexibility delivered by aggregators to third parties (i.e., DSOs, balancing capacity market).

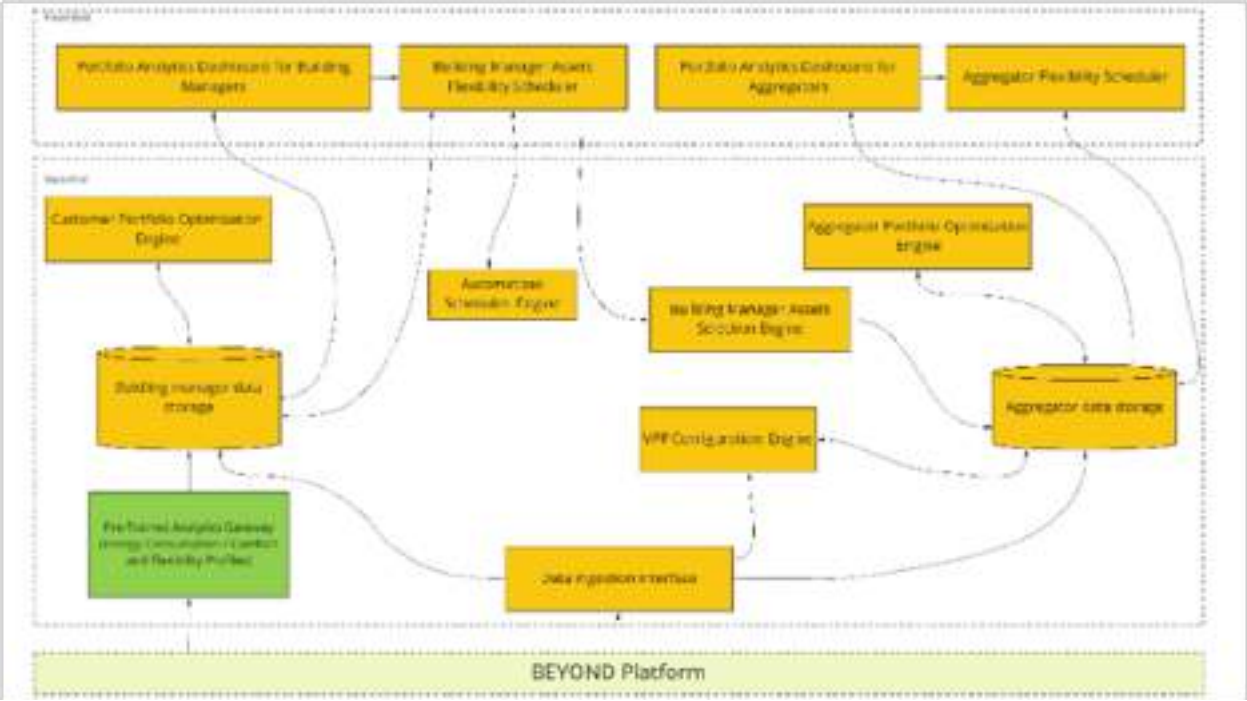

FIGURE 7-18: FLEXOPT TOOL ARCHITECTURE

7.9.4 Sequence Diagram

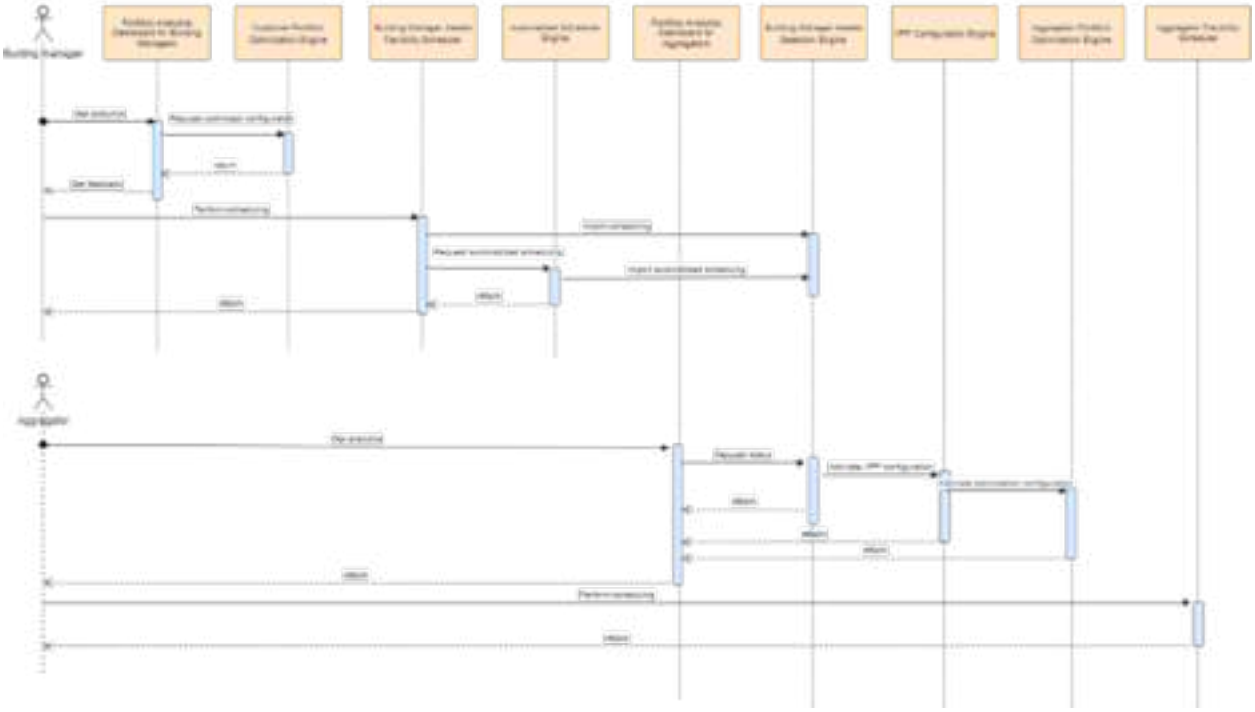

FIGURE 7-19: FLEXOPT SEQUENCE DIAGRAM

As shown in the sequence diagram, there are two types of users in the FLEXopt tool. The first one is the Building Manager which interacts with the Building Manager Dashboard to gain the insights about the building assets portfolio. Once this interaction is performed, the Dashboard communicates with the back-end and more specifically with the Customer Portfolio Optimisation Engine where optimisation algorithms are activated. Once the Building Manager has concluded his interaction with the Dashboard, it activates the Scheduler for the selection of flexibility to be delivered to the Aggregator.

The second user is the Aggregator who interacts with his own Dashboard and requests the insights of the available assets, through the Building Manager Assets Selection Engine. At the same time, once the Dashboard has been interacted, it activates the VPP Configuration Engine and the Optimisation Engine. Once all the insights have been collected, the Aggregator switches to the Flexibility Scheduler and based on the gain feedback from the analytics, schedules the flexibility to be delivered to third parties (i.e., DSO).

8 Conclusions & Next Steps

The deliverable at hand focuses on providing the first conceptual architecture and detailed description of the components, processes, and requirements of the BEYOND Platform. Moreover, it presents the platform's actors and their role in the different data-related processes as well as in the different end-user tools that were also described in detail to extract their functional and non-functional requirements.

Considering the conceptual architecture of the BEYOND Platform, an overview of its two layers namely the BEYOND Cloud based Platform and the BEYOND Private Infrastructure, as well as a brief description of their core components is presented. Furthermore, the role of each component and its interconnection with other components and sub-components is determined to reveal their provisions to the BEYOND Platform. It needs to be noted that for each core component that is documented, its technology stack and functional requirements are also reported accordingly. Along with the overall architecture and its underlying components, the BEYOND Platform's processes (i.e., data ingestion, data exploration, data trading, and data analysis) are presented and described through workflow diagrams (i.e., BPMN diagrams) providing a view on how the different components and sub-components interact to offer the different platform functionalities.

Regarding the end-user tools, provided by the different BEYOND business developers, an overview of their basic functionalities and architecture, as well as a description of their application's workflows, are determined. Additionally, the applications' functional and non-functional requirements are defined by the BEYOND business developers, while their corresponding mapping to the different use case and business requirements was defined accordingly.

Ultimately, this deliverable provides an initial version of the BEYOND Platform's overall conceptual architecture, its core elements, and the functional and non-functional requirements as extracted based on the different functionalities provided through the platform. This information along with the description of the end-user tools shall be adopted for future activities that involve any platform design and development within the context of WP3 "End-to-end Interoperable Big Data Management Platform", WP4 "AI Big Data Analytics Toolkit with Data Sharing Functions", and for future activities that involve the implementation of the different end-user tools within the context of WP5 "AI Analytics-based Decision Support Suite for Optimizing Energy Policy Planning, Infrastructure Sizing and De-risking Renovation Investments", and WP6 "AI Analytics-based Innovative Energy Services Suite towards Optimised Buildings Energy". According to the BEYOND DoA, T2.5 will remain active till M22, and thus the content of the deliverable at hand will be enriched and updated accordingly, while potential refinements obtained from BEYOND stakeholders' feedback and requirements, will be delivered through the final version of the BEYOND Platform architecture, namely the D2.7 "BEYOND Framework Architecture including functional, technical and communication specifications – b".

References

[1] BEYOND (2020) Description of Action (DoA)
[2] BEYOND (2021a): D2.1 - End-user & Business requirements analysis for big data-driven innovative energy services & ecosystems – a