Next Generation Meta Operating System

D1-Eros4NRG Detailed Design

Document Identification			
Status	Final	Due Date	24/05/2024
Version	1.0	Submission Date	23/05/2024

Related WP	WP1	Document Reference	D1
Related Deliverable(s)		Dissemination Level (*)	PU
Lead Participant	MARTEL INNOVATE B.V [MAR]	Lead Author	Panagiotis Kapsalis, Amjad Majid, Albert Seubers [MAR]
Contributors	Panagiotis Kapsalis, Amjad Majid, Albert Seubers [MAR]	Reviewers	MAR

Keywords:
Energy, IoT, Edge, Cloud, AI, ML, Analytics, One-Stop-Shop, Data, Enclave

Document Information

List of Contributors	
Name	Partner
Panagiotis Kapsalis	MAR
Amjad Majid	MAR
Albert Seubers	MAR

Document History			
Version	Date	Change editors	Changes
0.1	25/04/2024	MAR	ToC
0.2	30/04/2024	MAR	1st version of Section 1 and Section 2
0.3	02/05/2024	MAR	Updates on Section 1 and 2
0.4	10/05/2024	MAR	1st version of Section 3
0.5	14/05/2024	MAR	Updates on Section 3, 1st version of Section 4
0.6	17/05/2024	MAR	Updates on Section 4, Conclusions and next steps added
0.7	21/05/2024	MAR	Ready for Internal Review version
1.0	23/05/2024	MART	Final version

Table of Contents

List of Tables

List of Figures

List of Acronyms

Abbreviation / acronym	Description
AI	Artificial Inteligence
API	Application Programming Interface
D	Deliverable
DL	Deep Learning
DSO	Distribution System Operators
DR	Demand - Response
EC	European Commission
EU	Europe
EV	Electric Vehicle
FR	Functional Requirement
H2020	Horizon 2020
HE	Horizon Europe
HQ	Headquarter
IoT	Internet of Things
KER	Key Exploitable Result
ML	Machine Learning
MLOps	Machine Learning Operations
NFR	Non-Functional Requirements
NRG	Energy
PV	Photovoltaic
OSS	One Stop Shop
SW	Software
UC	Use Case
UI	User Interface
WP	Work Package

Executive Summary

Eros4NRG aims, through predictive analytics, to provide a platform for monitoring and analyzing energy-related assets. The Eros4NRG platform will serve as a One-Stop-Shop (OSS), enabling energy stakeholders to interact with their data and receive actionable predictions, thus enhancing decision-making processes within the energy domain.

The present document, **D1 – "Eros4NRG Detailed Design"**, provides the architectural specifications necessary to achieve this vision. The main outcomes reported in this deliverable include:

- **Definition of Eros4NRG Concept and Objectives:** Clearly outlines the core concept of Eros4NRG, identifies the target stakeholders, addresses the challenges to be tackled, and specifies the main objectives of the project.
- **Analysis of Reference Architectures:** Examines notable reference architectures that focus on developing IoT/Edge/Cloud systems within the energy sector, offering insights that inform Eros4NRG's architectural design.
- **Use Cases, Non-Functional, and Functional Requirements:** Presents detailed use cases, identifies both non-functional and functional requirements, and describes the system architecture and its building blocks.
- **API, Plugin, and User Interface Specifications:** Provides comprehensive specifications for the APIs, plugins, and user interfaces that will be part of the Eros4NRG platform.
- **Economic and Business Impact:** Defines the potential economic and business impact of Eros4NRG, emphasizing its relevance and value to the energy market.

The information and specifications provided in this deliverable are designed to guide the development, integration, and validation activities of the Eros4NRG project, ensuring a structured approach to achieving its ambitious goals.

1 Motivation and Technological Background of the proposed IoT Application

Eros4NRG is envisioned to be a comprehensive platform for monitoring and analysing energy production and consumption. Through data-driven predictive analytics, Eros4NRG enables end users to make informed decisions about their energy operations, leading to more efficient business operations. This platform will also serve as a one-stop shop (OSS) for energy stakeholders. To realise this vision, we plan to involve in the development process two main stakeholder groups:

- **Electromobility Managers** (EMOTION), main entities that produce the data assets which will be exploited by Eros4NRG are EVs (Electric Vehicles) and EV charging stations.
- **DSO** (ASM Terni), main entities that produce data that will be exploited and later processed by Eros4NRG are PVs (photovoltaic plants) and ASM Smart HQs.

In addition to enegy data analytics, Eros4NRG aims to enhance data trust & transparency by analysing and securing IoT and edge data. Also, establishing cutting-edge data storage solutions that enhance semantic connections between dynamic information and metadata. Finally, Eros4NRG aims to enhance the Nemo Project Smart Energy Trial[1] by acting as a native extention to it. To realize this vision, Eros4NRG will address the following key challenges:

- **Unreliable Energy IoT/Data Sources:** Addressing the impact of Inaccurate data in energy ecosystems (Buildings, PVs, EVs) on ML operations and training.
- **Lack of Organization and Management of Energy Data:** Tackling overlooked static data elements like energy entity attributes, environment, geolocation, and device fleets, leading to information loss and diminished analytics quality.
- **Involving Stakeholders in Cross-Value Chain Services:** Enhancing AI/ML model explainability and translating machine learning predictions into user-friendly language to make them comprehensible to non-technical stakeholders.

The objectives of Eros4NRG are to:

- **Increase Data Trust and Transparency in Energy Analytics Services:** Conduct data inspection and anomaly detection to identify anomalous data sources.
- **Improve Data Self-Explainability, Management and Secure Sharing:** By enhancing data management via metadata extraction, which improves semantic connections and fosters a secure and accessible data ecosystem.
- **Deploy High-Quality Energy Analytics:** Utilizing high-quality checked-in data to train AI/ML models to optimize energy grid behavior.
- **Introduce AI/ML Models Explainability:** Implementing explainability components that enable end-users to make more informed, energy-efficient decisions.
- **Deliver a one-stop-shop (OSS) for predictive energy analytics**: An OSS has the potential to be a centralised platform that offers all the necessary tools and resources for utilising predictive analytics in the energy sector. This platform will include features such as data collection, analysis, modelling, and forecasting, as well as tools for visualisation and interpretation of results.

1.1 Background: EU-funded IoT projects in the energy sector

In this section, we review several pivotal IoT-related EU-funded projects in the energy sector, tracing the critical processes from data acquisition and enrichment through to processing, knowledge extraction, and ultimately, monetization.

[1] https://meta-os.eu/index.php/trial-3/

1.1.1 MATRYCS (H2020)

The MATRYCS[2] aims to capitalise and combine existing modern technological breakthroughs in the areas of ML / DL and big data, in order to develop a new decision-making and data analytics solution for energy-efficient buildings. MATRYCS will realise a holistic, state-of-the-art AI-empowered framework for decision-support models, data analytics and visualisations for Digital Building Twins and real-life applications. The overall vision of MATRYCS is to define and deploy a Reference Architecture[3] for Buildings Data exchange, management and real-time processing, and to translate this reference architecture into an Open, Cloud-based Data Analytics Toolbox (MATRYCS Modular Toolbox). It will enable AI-based cross-sector analytics for smart energy-efficient buildings through three layers, MATRYCS-GOVERNANCE, MATRYCS-PROCESSING and MATRYCS-ANALYTICS. Eventually, MATRYCS will enable reliable and effective policymaking, as well as support the creation and exploitation of innovative services through the utilization of a wide variety of data, for the safe and effective operation of buildings. MATRYCS service analytics reference framework will be applied, demonstrated and validated in 11 real-life large-scale pilots strategically selected to have multinational direction covering different regions and levels, such as regional, national and pan-European.

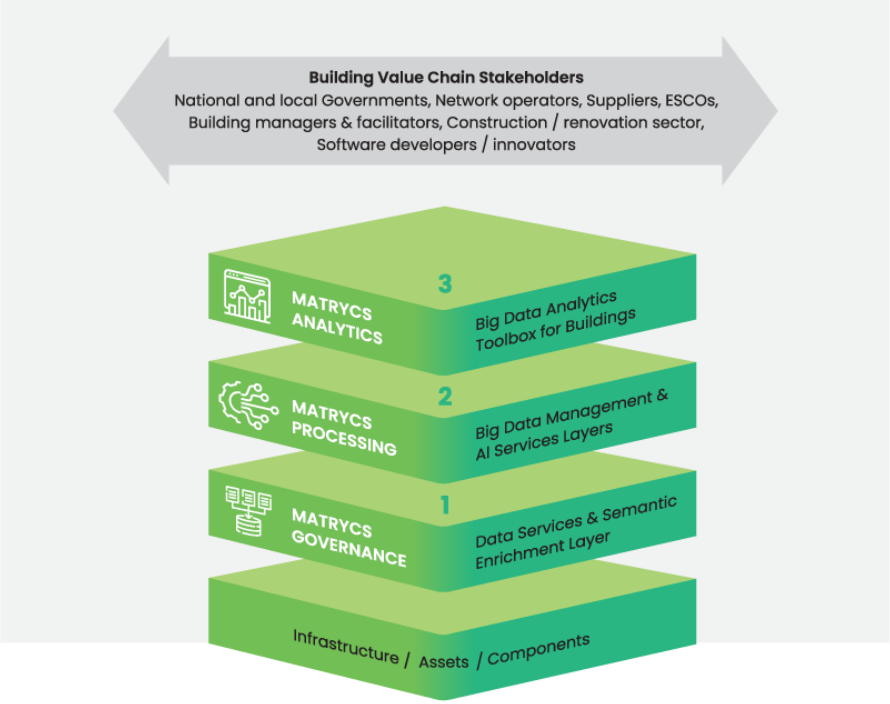

Figure 1: MATRYCS H2020 Reference Architecture

[2] https://matrycs.eu/

[3] Pau, M., Kapsalis, P., Pan, Z., Korbakis, G., Pellegrino, D., & Monti, A. (2022). MATRYCS—A big data architecture for advanced services in the building domain. Energies, 15(7), 2568.

1.1.2 DigiBUILD (HE)

DigiBUILD[4] will catalyse this much-needed transformation by making use of high-quality data and next generation digital building services, supporting the deployment of EU-wide Framework for a Digital Building Logbook. An inclusive environment for multi-stakeholder knowledge exchange (based on European Bauhaus initiative) will be applied to co-design end-user-oriented services. DigiBUILD will provide an open, interoperable, and cloud-based toolbox[5] to transform current 'silo' buildings into digital, interoperable and smarter ones, based on consistent and reliable data, supporting better-informed decision-making for performance monitoring & assessment, planning of building infrastructure, policy making and de-risking investments. It will be built on top of existing platforms and common EU initiatives, towards an Energy Efficient Building Data Space, based on standard cloud-data platform frameworks (FIWARE) and Data Space initiatives (GAIA-X and IDSA). On top of this advanced data governance framework, we will create AI-based data analytics and Digital Building Twins based on high-quality data, aiming to facilitate transparency, trust, informed decision-making and information sharing within the built environment and construction sector, which will be deployed across 10 real-world conditions (TRL 8).

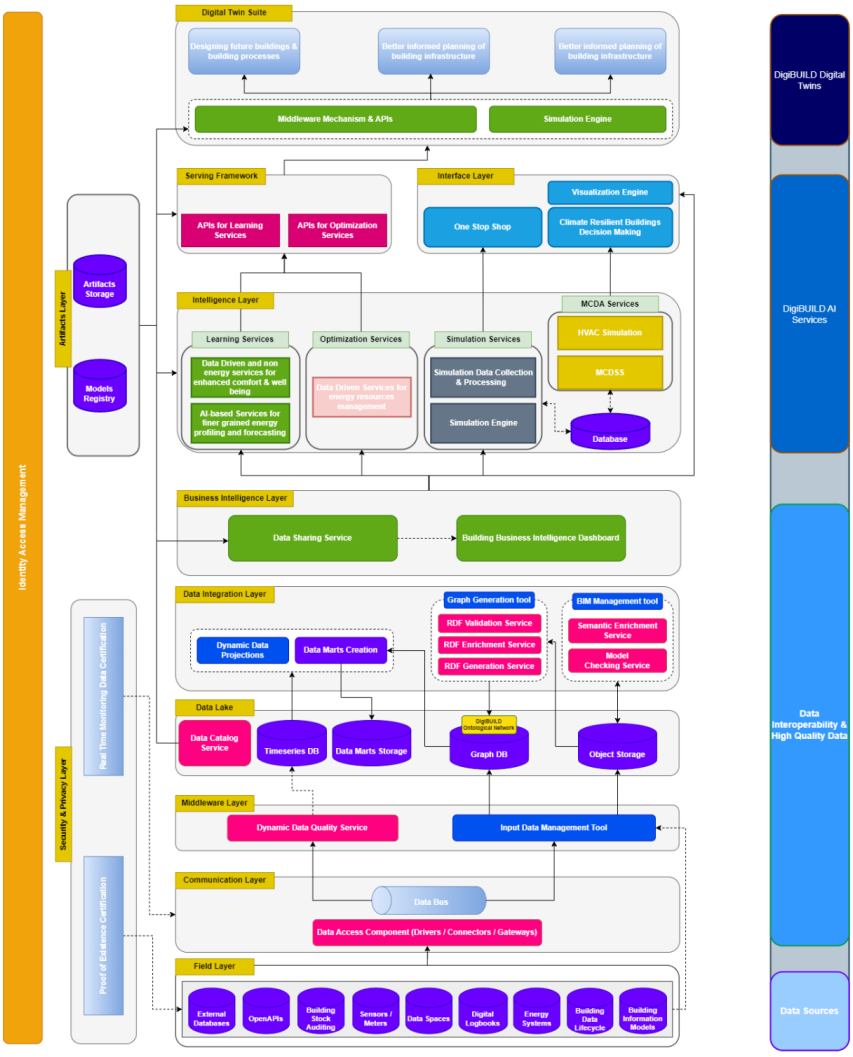

Figure 1: DigiBUILD Reference Architecture

[4] https://digibuild-project.eu/

[5] Hernández, J. L., Martín, S., Kapsalis, P., Katsigarakis, K., Sarmas, E., & Marinakis, V. (2023, July). Building a Data Lake for Smart Building Data: Architecture for data quality and interoperability. In 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA) (pp. 1-8). IEEE.

1.1.3 BEYOND (H2020)

BEYOND H2020[6] project develops and offers a big data platform and a set of technologies that allow a data consumer to search, find and utilize data generated by buildings (data owners). Based on these, the data consumers can run analytics and simulations that are actually needed to design a project and exploit them during the real-time operation of the buildings so as to optimize their operation and energy performance. Beyond that, the whole process is concluded between the data providers and data consumers under a data smart contract[7] that allows a data provider to get remunerated. On the other hand, the data consumer can satisfy its business functions while at the same time can run analytics on this data and gradually, turn into a data provider itself by making available these insights to other stakeholders in the energy market. As a result, the BEYOND H2020 project is giving the opportunity to these actors to trade the data analytics to data consumers so that the latest evaluate to what degree the buildings can generate revenues through their participation in emerging trends in the energy ecosystem like flexibility trading. Thus, the BEYOND H2020 project nurtures an ecosystem that is not only focused on services but in data exchange, and most importantly, on the monetization of the buildings' data.

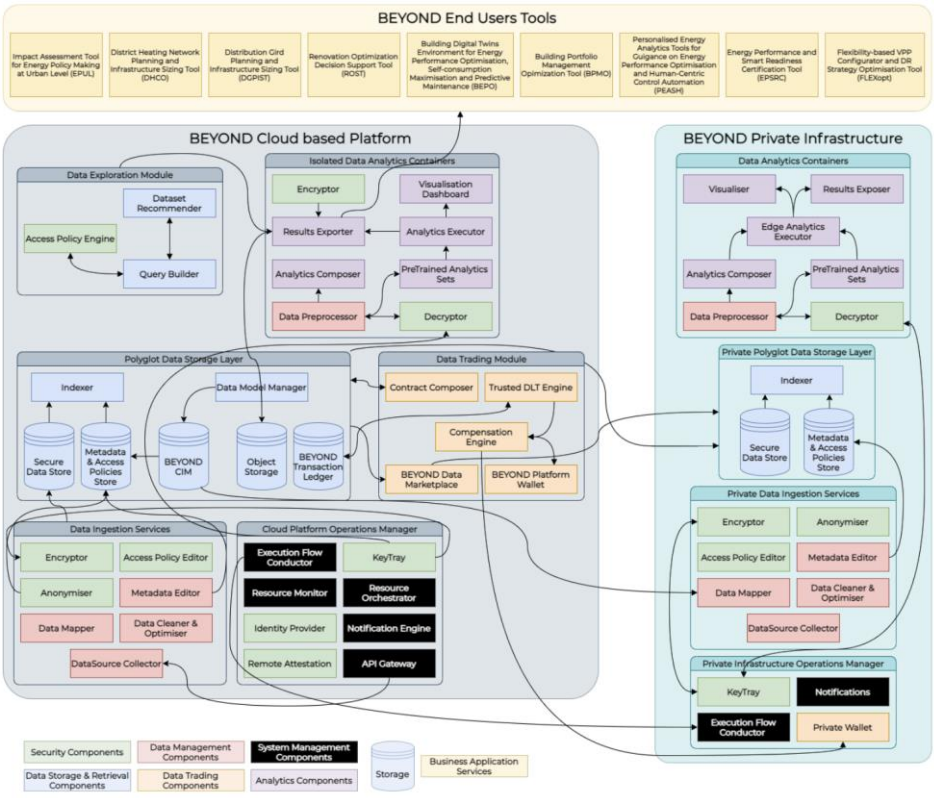

Figure 2: BEYOND H2020 Reference Architecture

[6] https://beyond-h2020.eu/

[7] Foti, Magda. "Privacy-Preserving Market-Driven Transactive Energy System Using Homomorphic Encryption." 2023 19th International Conference on the European Energy Market (EEM). IEEE, 2023.

2 Detailed Design of the SW Components/Plugins including Functional and Non-Functional Requirements

This chapter aims to illustrate the design of the Eros4NRG Software (SW) components and plugins. It includes a detailed overview of both the non-functional and functional requirements that will shape the Eros4NRG system architecture. The methodology employed comprises several key steps. Initially, interviews and workshops were conducted with ASM and EMOTION, who are the intended end users of the project. The insights gained from these sessions established two critical use cases for Eros4NRG. Following this, the non-functional requirements were identified and rigorously analyzed. Building on this, the functional requirements were then elicited, ensuring a comprehensive and coherent framework. The chapter concludes by outlining the Eros4NRG system architecture, which integrates all the elements discussed into a unified whole. Figure 4 illustrates the design methodology for Eros4NRG.

Figure 3: Eros4NRG Design Process

2.1 UC.1 – Forecasting Energy Assets to support ASM Terni Demand-Response Strategies

ASM Terni utilizes a variety of IoT/Edge components, such as Power Quality Analyzers, to measure and monitor the energy consumption at ASM headquarters and the energy production from photovoltaic (PV) systems. As a key stakeholder of Eros4NRG, ASM expects the Eros4NRG One-Stop-Shop (OSS) to provide comprehensive real-time analytics and visualizations. These tools are designed to enable end users to monitor and manage the ASM energy systems more effectively. A crucial output from this platform is the accurate forecasting of PV production and headquarters' energy consumption. Such forecasts are essential for optimizing ASM's energy distribution and effectively managing demand-response strategies. This not only aids in better managing the energy supply from renewable sources but also reduces operational costs and enhances the overall efficiency of the Terni energy grid. Moreover, a user-friendly interface with easy-to-interpret visualizations of energy flows and consumption patterns would add value by revealing hidden patterns, thus assisting end users in decision-making processes. Table 1 outlines the Non-Functional Requirements (NFRs) related to UC1.

Table 1: Non-Functional Requirements for UC1

NFR ID	Description
NFR.1.1	Capture data generated by ASM Terni IoT/Edge devices and provide a Data Lake for hosting extracted raw information.
NFR.1.2	Identify and exclude abnormal data to ensure that only reliable, high-quality information is retained.
NFR.1.3	Pre-process raw data, anonymize sensitive information, and homogenize the clean data for further use.

NFR.1.4	Store processed data to facilitate enhanced querying and AI/ML training capabilities
NFR.1.5	Provide three-hour advanced forecasts for PV energy production and ASM HQs energy consumption
NFR.1.6	Deliver advanced visualizations and reports for ASM curated and hosted data in Eros4NRG, assisting ASM stakeholders in making informed decisions about their energy infrastructure behavior.

2.2 UC.2 – Continuous Monitoring of EMOTION EVs State of Charge

EMOTION operates three EV charging stations and maintains a fleet of six electric vehicles (EVs). Currently, the EMOTION systems lack crucial functionalities necessary for electromobility managers (EMOTION stakeholders) to effectively utilize and analyze both historical and real-time data generated by the EV batteries and charging stations. Eros4NRG aims to provide EMOTION stakeholders with comprehensive analytics, both visual and predictive. This includes insights into the utilization of charging stations over time, detailing the charging sessions conducted at each station and plug, with a specific emphasis on time windows. Additionally, Eros4NRG will provide detailed information on EVs, particularly regarding their state of charge, and predictive insights about the expected lifespan of EV batteries. These enhancements will empower EMOTION stakeholders to make informed decisions and optimize the efficiency and longevity of their EV fleet. Table 2 outlines the Non-Functional Requirements (NFRs) related to UC2.

Table 2: Non-Functional Requirements for UC2

NFR ID	Description
NFR.2.1	Capture time series data representing various aspects of EV operation, including battery status, velocity, number of charging sessions, total kilometers traveled, and kilowatt-hours charged.
NFR.2.2	Capture data related to EV charging sessions, including the state of plugs and kilowatt-hours provided during each session.
NFR.2.3	Pre-process raw information to anonymize sensitive information in charging sessions and in EVs state of charge (users' related information)
NFR.2.4	Store and organize extracted information to perform effective querying of historical and real time data related to EV charging sessions and EV battery status.
NFR.2.5	Provide visualizations and reports on historical and real-time data about EV charging stations and the state of EV batteries. Perform queries to gain insights on power production over time at EVs and stations, and retrieve insights about the number of charging sessions performed at each station, elaborating on each plug within selected time windows.
NFR.2.6	Retrieve forecasts (1 hour ahead) for the state of EV batteries.
NFR.2.7	Retrieve forecasts (3 hours ahead) regarding charging session duration.

2.3 Eros4NRG Functional Requirements

Eros4NRG functional requirements are based on the non-functional requirements listed in Table 1 and 2. They are used to define the Eros4NRG's system architecture.

Table 3: Eros4NRG Functional Requirements

FR ID	Short Description	Relevant NFRs
FR.001	Mechanisms to capture data generated by target IoT/Edge Entities	NFR.1.1, NFR.2.1, NFR.2.2
FR.002	Provide state-of-the-art anomaly detection techniques to identify abnormal data, ensuring that only high-quality information is retained. To achieve this a dedicated buckets organisation within the data lake will host the raw captured information into the "raw bucket" and normal information, checked-in by anomaly detection to the "clean bucket".	NFR.1.1, NFR.1.2, NFR.2.1, NFR.2.2, NFR.2.3
FR.003	Data pipelines will run on top the clean bucket to refine the checked-in information. Specifically, processes such as curation, anonymization, harmonisation, and data modelling will be applied to prepare the data for efficient querying and AI/ML training.	NFR.1.3, NFR.2.3
FR.004	Eros4NRG Data Enclaves will play a vital role in enabling advanced querying capabilities by harnessing the potential of photovoltaic systems (PVs) and smart building attributes. These Enclaves serve as the foundation for shaping the ASM Terni metadata model, known as the Graph Enclave, drawing upon open-source schemas such as FIWARE Smart Buildings, HAYSTACK, SAREF, and BRICK. Within these Enclaves, time-series information, such as power production from PVs and energy consumption at ASM HQ, is dynamically stored in the data enclave. Establishing a semantic link between these two Enclaves is crucial for fostering attribute-based querying and facilitating seamless information processing.	NFR.1.3, NFR.1.4
FR.005	Eros4NRG Data Enclaves will play a pivotal role in enabling advanced querying capabilities to retrieve real time data from EVs and charging stations.	NFR.2.3, NFR.2.4
FR.006	Eros4NRG will integrate Workflow Orchestration mechanisms (e.g. Prefect, Apache Airflow) to optimize data and AI/ML pipelines and orchestrate their execution in Cloud Native Infrastructure (K8S, microK8S, Docker)	NFR.1.1, NFR.1.2, NFR.1.3, NFR.1.4, NFR.1.5, NFR.2.1, NFR.2.2, NFR.2.3, NFR.2.4, NFR.2.5, NFR.2.6
FR.007	AI/ML models for PV energy production forecasting and ASM HQ energy consumption forecasting will be implemented, using different techniques (Neural Models, Statistical Analysis, N-BEATS, Temporal Convolutional Networks) and benchmark the models using different lags and horizons.	NFR.1.5
FR.008	Eros4NRG will provide a Secure Data Catalogue to export stored information, data assets in Eros4NRG Data Enclaves.	NFR.1.4, NFR.2.4

FR.009	Eros4NRG platform to provide AI/ML driven inferences to predict the future state of EVs' batteries and the charging session duration. Eros4NRG platform aims to implement models using different techniques like ANNs, LSTMs, RNNs, N-BEATS, TCNs and state-of-the libraries like Scikit-Learn, XGBoost, DARTS, Keras, Tensorflow.	NFR.2.6, NFR.2.7
FR.010	Eros4NRG will integrate NEMO Unified Access Control to integrate OAuth2.0 and UMA 2.0 protocols which will enable role-based and resource-based access control over Eros4NRG components like Data Catalogue, and OSS.	NFR.1.4, NFR.2.4, NFR. 1.6, NFR.2.5
FR.011	Eros4NRG will integrate Nemo CyberSecure & Federated MLops mechanism to store version and serve trained AI/ML models.	NFR.1.5, NFR.1.6, NFR.2.6, NFR.2.7
FR.012	Eros4NRG OSS will offer real-time visualisations of the Collaborative data enclaves for ASM, effectively representing ASM's data assets and patterns in a user-friendly manner. Additionally, the OSS will deliver forecast results and time-series analysis for PV energy production and ASM HQ energy consumption, enhancing decision-making and strategic planning.	NFR. 1.6, NFR.2.5

2.4 Eros4NRG System Architecture and Core Components

Eros4NRG relies on a comprehensive set of both functional and non-functional requirements to shape its system architecture. This architecture is designed to be service-oriented and seamlessly compatible with modern cloud-native infrastructures such as Kubernetes, MicroK8S, and Docker. At the heart of the Eros4NRG platform lie its core components, which include:

- **Zero Trust Data Services**: This component plays a pivotal role in Eros4NRG's operations by extracting crucial information from ASM Terni and EMOTION IoT/Edge Fleets. Its duties encompass not only pre-processing tasks but also metadata decomposition, ensuring that data is appropriately structured and prepared for further analysis.
- **Collaborative Data Enclaves**: This section houses the storage components and data models that harmonize with the data assets.
- **One-Stop-Shop for Energy (OSS 4 Energy):** This upstream component serves as a dashboard, showcasing the results of AI/ML models that deliver predictive analytics within the energy sector. It also facilitates advanced visualizations, reports, and queries for enhanced insight and decision-making.

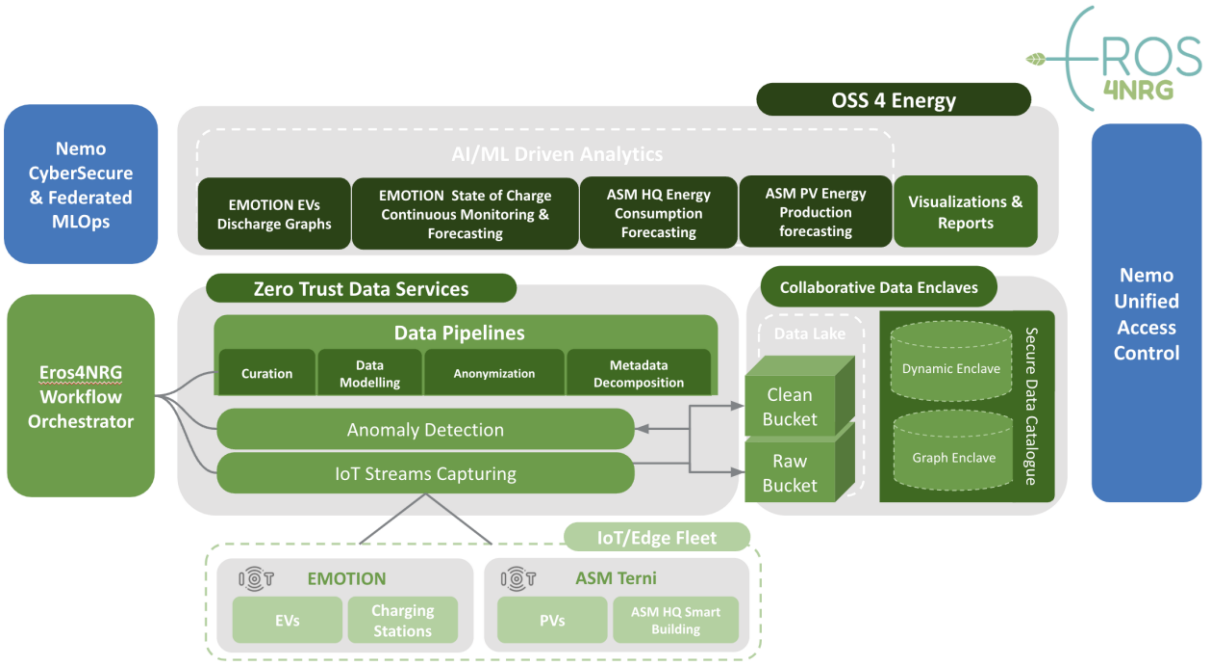

Figure 4: Eros4NRG System Architecture

Eros4NRG is integrated with **NEMO CyberSecure & Federated MLOps** to store, version and serve trained models and **NEMO Unified Access Control** to provide role-based and resource-based access control on top of Eros4NRG assets (data, models and charts).

Table 4: Integration with NEMO

NEMO COMPONENT	Functionality to cover	Integration with Eros4NRG
NEMO CyberSecure & Federated MLOps	Provide models keeping, versioning and serving capabilities for Eros4NRG AI/ML Models.	EROS4NRG AI/ML Driven Analytics
NEMO Unified Access Control	Enable role-based access control in Eros4NRG. Use OAuth2.0 access tokens.	Secure Data Catalogue; Visualisations and Reports

2.4.1 Zero Trust Data Services

This component is responsible for the following functionalities:

- **IoT Streams Capturing**: Capture IoT streams generated by EMOTION and ASM Terni IoT/Edge devices.
- **Anomaly Detection**: Identify anomalies within captured data, isolate the anomalous data, and store normal information in designated clean buckets, ensuring that checked-in data remains pristine.
- **Workflow Orchestrator**: Deploy and initiate data jobs and pipelines as microservices, with each task, job, or pipeline orchestrated as a container within the underlying cloud-native infrastructure.
- **Data Pipelines**: The data pipelines are tasked with operating on pristine data, carrying out essential pre-processing routines such as curation, data modeling, anonymization, and metadata decomposition. These pipelines are executed as microservices within the cloud-native infrastructure, harnessing the capabilities of Eros4NRG Workflow Orchestration for seamless orchestration and management.

The following picture demonstrates the interactions between Zero Trust Data Services modules. As mentioned before key functionality of Zero Trust Data Services is the metadata decomposition where static information is detached from timeseries information (dynamic information). Static information is modelled according to Open Standards and Open models like FIWARE[8], HAYSTACK[9] and SAREF[10] and the generated data streams will be stored in timeseries databases.

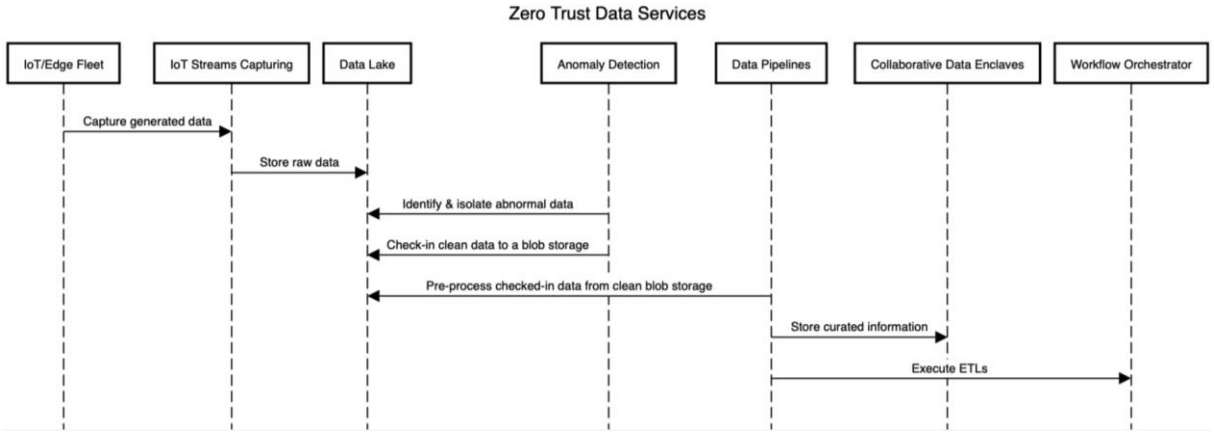

Figure 5: Zero Trust Data Services Interaction Diagram

The following table includes libraries and technologies that will be used for the technical implementation of the Zero Trust Data Services.

Table 5: Libraries/Technologies that will be used for Zero Trust Data Services Development

Module	Short Description	Integration with Other Components / Layers	Libraries/Technologies
IoT Streams Capturing	Captures data from EMOTION and ASM IoT/Edge devices and store them to the Data Lake (raw).	IoT/Edge Fleet; Collaborative Data Enclaves	Python 3.10[11], Python requests[12], Pandas[13].
Anomaly Detection	Retrieve data from the raw storage (Data Lake) perform anomaly detection and stores clean data back to the data lake to a "clean object"	Collaborative Data Enclaves	Scikit-Learn[14], Keras[15], Tensorflow[16], Pandas
Data Pipelines	ETLs to perform curation, anonymization, semantic enrichment and stores the results to Data Enclaves	Collaborative Data Enclaves	Pandas, Scikit-Learn

[8] https://fiware.org/smart-data-models/
[9] https://www.project-haystack.org/
[10] https://saref.etsi.org/
[11] https://www.python.org/
[12] https://pypi.org/project/requests/
[13] https://pandas.pydata.org/
[14] https://scikit-learn.org/stable/
[15] https://keras.io/
[16] https://www.tensorflow.org/

Workflow Orchestrator	Triggers/Executes the Data Pipelines as pods/containers to the Cloud Native Infrastructure.	Data Pipelines	K8S[17], MicroK8S[18], Prefect[19], Apache Airflow, Celery[20], Dask[21]

2.4.2 Collaborative Data Enclaves

Collaborative Data Enclaves are hosting the storage components and in particular:

- **Data Lake**: This component serves as a blob storage, with a "Raw bucket" housing data immediately after capture, and a "Clean Bucket" dedicated solely to verified, pristine data.
- **Dynamic Enclave**: These dynamic repositories serve as hosts for the time-series information generated by the IoT/Edge layer. The stored data is formatted as [(date, value)], providing a structured representation of temporal data points.
- **Graph Enclave**: This graph database serves as a repository for static information, leveraging interconnected entities to enhance query efficiency by uncovering hidden insights. It includes semantic links to the Dynamic Enclave, facilitating connections between data entities. For example, the time series generated by a sensor located on the 1st floor of a specific building correlates to a specific table within the Dynamic Enclave.
- **Secure Data Catalogue**: This component is the set of APIs and Data exporters which export the stored information in Dynamic and Graph Enclaves. The semantic links between these two types of enclaves (collaboration) is the connection which connects the physical entities and attributes with dynamic information. The Data Catalogue will be protected using the capabilities of NEMO Unified Access Control component to apply role-based access control and restrict the access over the Eros4NRG data assets.

Table 6: Technologies/Libraries that will be used for the Collabotive Data Enclaves Development

Module	Short Description	Integration with Other Components / Layers	Libraries/Technologies
Data Lake	Is the blob storage configured to operate on top of a Cloud Native Infrastructure and will host Raw and Clean buckets (directories) after data check-in.	Zero Trust Data Services	MinIO[22]
Graph Enclave	Will be the graph database that will host the Eros4NRG data model and metadata. Will also contain as properties the connections of the metadata with dynamic information (timeseries).	Zero Trust Data Services	Neo4j[23]

[17] https://kubernetes.io/
[18] https://microk8s.io/
[19] https://www.prefect.io/
[20] https://docs.celeryq.dev/
[21] https://www.dask.org/
[22] https://min.io/
[23] https://neo4j.com/

Dynamic Enclave	Will be the timeseries database that will host the timeseries information generated by ASM Terni and EMOTION IoT/Edge devices	Zero Trust Data Services	TimescaleDB[24]
Secure Data Catalogue	Will be the APIs and data exporters responsible to query the data enclaves (graph db and timeseries) and export that information. Will integrate NEMO Unified Access Control (Keycloak) component to apply role-based access control over the data assets.	Zero Trust Data Services; One-Sto-Shop for Energy; NEMO Unified Access Control	Python 3.10, FastAPI[25], Pandas, Docker, Kubernetes

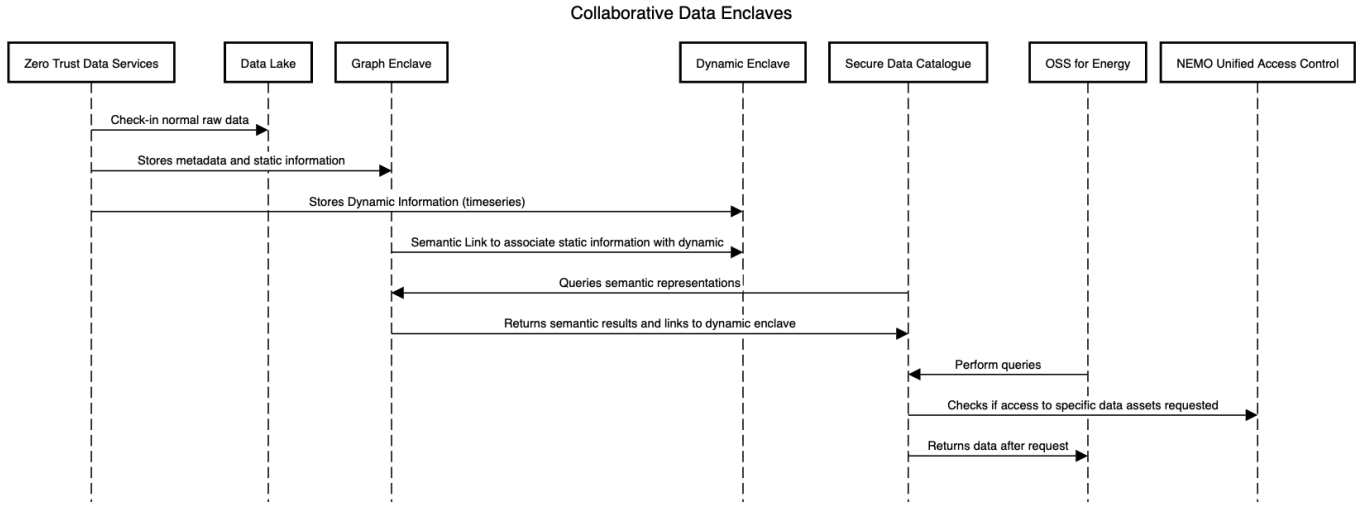

Figure 6: Collaborative Data Enclaves Interaction Diagram

2.4.3 One-Stop-Shop for Energy

Eros4NRG One-Stop-Shop (OSS) for Energy will be an application which will offer the following capabilities to its end users (ASM Terni, EMOTION):

- Deliver insights through visual representations.
- Provide data reports, advanced querying functionalities leveraging energy-related metadata.
- Advanced end users will have the capability to query the Eros4NRG data infrastructure and create their own visualizations and charts.
- Deliver state-of-the art trained models tailored made for ASM Terni and EMOTION UCs and in particular:
  - (**EMOTION**): EVs state of charge continuous monitoring & forecasting, EVs discharge graphs
  - (**ASM Terni**): PVs energy production forecasting, ASM HQs energy consumption forecasting
- Provide capabilities to store, version, and provide access to trained models for end users, by integrating NEMO CyberSecure MLOps.

[24] https://www.timescale.com/
[25] https://fastapi.tiangolo.com/

- Eros4NRG will integrate NEMO Unified Access Control to secure all its assets, including data, models, visualizations, and queries. For example, only EMOTION end users will be permitted to interact with EMOTION-related assets, and similarly, only ASM Terni end users will have access to ASM Terni-related assets.

The main components included in Eros4NRG OSS are:

- **Visualizations & Reports Engine**: Will be a Business Intelligence Engine where visualizations and reports will be provided to the end users. Will integrate NEMO Unified Access Control to apply role-based access control to restict access.
- **Eros4NRG AI/ML Driven Analytics**: A set of models will leverage the Eros4NRG Secure Data Catalogue to train AI/ML models for ASM and EMOTION. These models will be stored and versioned in the NEMO CyberSecure MLOps component. The results, metrics, and predictions from these models will be visualized as reports and charts in the Visualizations & Reports Engine.

Eros4NRG OSS functionalities include 3 main flows.

The **1st flow (Visualizations/Queries flow)** illustrates the process by which users can generate and access visualizations and queries within the OSS (One-Stop-Shop), highlighting the interaction between various components. Specifically, users are guided through the retrieval or creation of specific visualizations, reports, and queries within the Visualization & Reports Engine. The NEMO Unified Access Control system intercepts users' access tokens, assessing their roles (whether they belong to EMOTION or ASM Terni organization), and grants or denies access to the requested resources accordingly. Secure Data Catalogue serves as the gateway to data for the Visualization and Reports Engine, ensuring secure and authorized access.

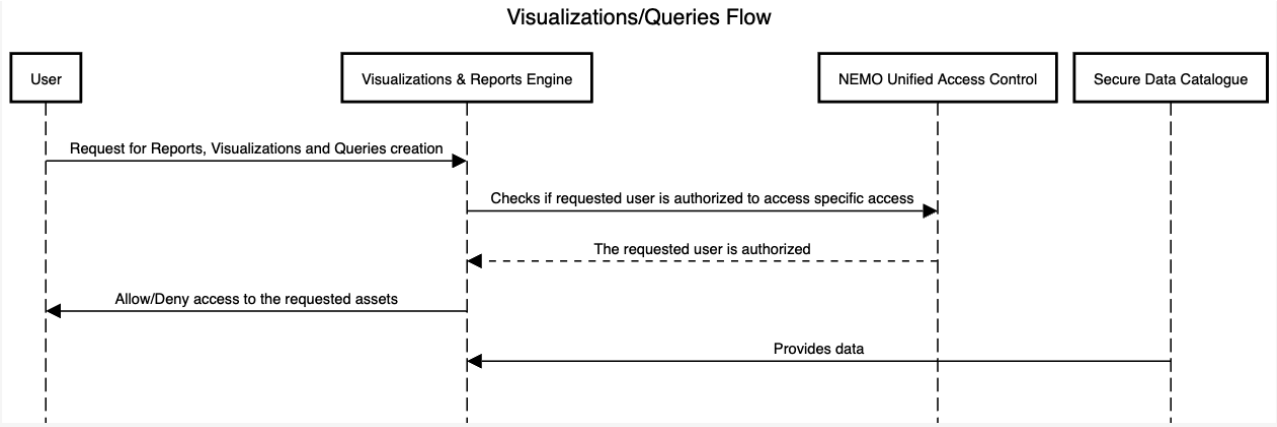

Figure 7: Visualizations/Queries Flow in OSS

The **2nd flow (OSS AI/ML Training flow)** depicts the AI/ML training process and the interactions between Eros4NRG and NEMO Components. The Secure Data Catalogue will supply data to the AI/ML Driven Analytics module, where AI/ML pipelines will be executed as pods or containers within a cloud-native infrastructure. This setup will utilize the capabilities of the Workflow Orchestrator. Additionally, the Workflow Orchestrator will schedule retraining jobs to keep the models up to date. Once the AI/ML training is complete, the training artifacts and metadata will be stored in NEMO CyberSecure MLOps components. These results will also be presented to the end users as reports and charts by the Visualizations & Reports Engine.

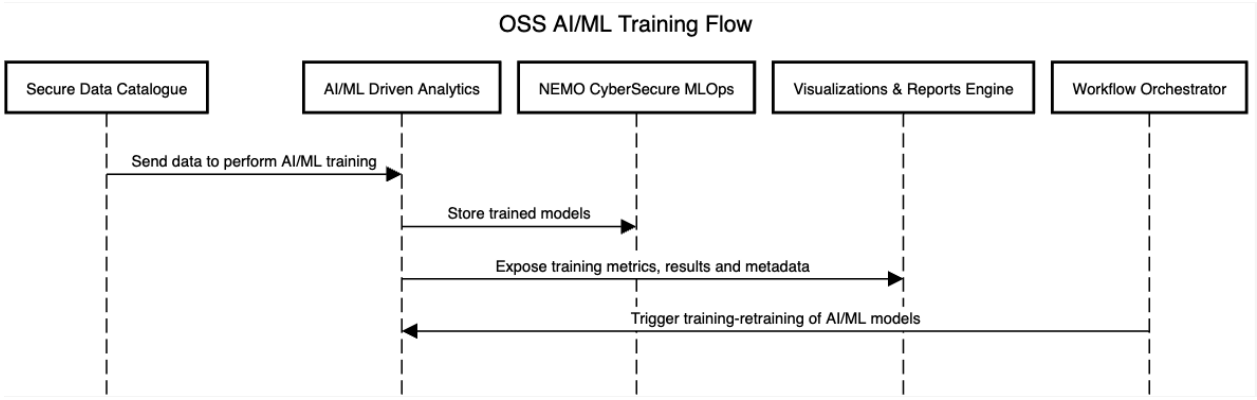

Figure 8: AI/ML Training Flow in OSS

The **3rd flow (OSS Prediction Flow)** demonstrates the process of making predictions with new data (data not used during training procedure). When an end user requests predictions from the Visualizations & Reports Engine, it utilizes models trained in the second workflow for a specified time window. The access token is introspected to ensure the user has permission to access the relevant data and models. If the user has access, the Secure Data Catalogue selects the appropriate time window and provides this as new/test data for prediction. The trained models are retrieved from NEMO CyberSecure MLOps, and the predictions are then delivered to the end user via REST APIs or displayed in the Visualizations & Reports Engine.

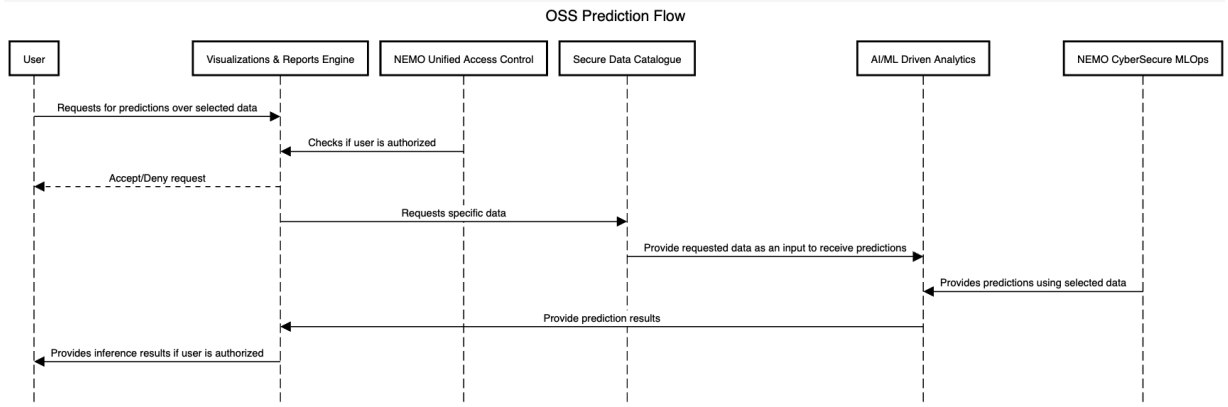

Figure 9: OSS Prediction Flow

Table 7: Technologies/Libraries that will be used for Eros4NRG OSS development

Module	Short Description	Integration with Other Components / Layers	Libraries/Technologies
Visualizations & Reports Engine	Provides visualizations, reports and charts over stored information and metrics, training artifacts from the AI/ML driven analytics	NEMO Unified Access Control; NEMO CyberSecure MLOps; Secure Data Catalogue; AI/ML Driven Analytics;	Apache Superset[26]; PostgreSQL[27]; Django[28]

[26] https://superset.apache.org/
[27] https://www.postgresql.org/
[28] https://www.djangoproject.com/

AI/ML Driven Analytics	AI/ML pipelines responsible for training/retraining of AI/ML models for energy consumption/production forecasting (ASM Terni), EVs batteries state of charge and Session Duration (EMOTION)	NEMO CyberSecure MLOps; Visualization & Reports Engine; Workflow Orchestrator; Secure Data Catalogue	Keras[29]; Tensorflow[30]; Scikit-Learn

[29] https://keras.io/
[30] https://www.tensorflow.org/

3 Detailed API of the SW Components Plugins

3.1 Zero Trust Data Services APIs and Plugins

Table 8: API Endpoint – IoT Streams Capturing

Trigger data capturing from ASM Terni and EMOTION IoT/Edge Devices	
Description	Capture ASM Terni and EMOTION Data and store them to Data Lake (raw bucket)
HTTP Method	GET
Endpoint	`/capture?source=${emotion | asm}`
Parameters	Query String: select value for source to extract data from emotion or/and asm IoT/Edge devices
Output	HTTP Response Code: 200

Table 9: Plugin - Anomaly Detection

Anomaly Detection Plugin	
Description	Plugin/Daemon to schedule Anomaly Detection over the raw bucket in Eros4NRG Data Lake. It schedules periodical anomaly detection tasks to retrieve raw data and store/check in normal data to a new clean bucket.
Inputs	Load raw data from Data Lake

Example	`import minio`
	`from celery import Celery`
	
	`app = Celery('anomaly_detection')`
	`app.config_from_object('settings', namespace='CELERY_')`
	
	`@app.task()`
	`def load_raw_data():`
	`    """This task will load raw data"""`
	`@app.task()`
	`def select_normal_data():`
	`    """This task will select only normal data"""`
	`@app.task()`
	`def store_clean_data():`
	`    """This task will store clean/normal data to a new bucket in MinIO"""`

Outputs	Store normal/clean data to a new bucket (clean bucket) in Data Lake.

Table 10: Plugin – Data Pipelines in Workflow Orchestrator

Workflow Orchestrator	
Description	Integrates data and AI/ML pipelines and executes them as pods/containers to Cloud Native Infrastructure (K8S).

Inputs	Data Pipelines; AI/ML Training Pipelines; Python Scripts

Example	`from prefect import flow, get_run_logger, task, get_run_logger`
	`from prefect import variables`
	`@task`
	`def curation(data):`
	`  return curated_data`
	`@task`
	`def anonymization(curated_data):`
	`  return anonymized_data`
	`@task`
	`def metadata_decomposition(anonymized_data):`
	`  return static_data, timeseries, semantic_links`
	`@task`
	`def store_to_data_enclaves(static_data, timeseries, semantic_links):`
	`@flow`
	`def pipeline():`
	`  curated_data = curation()`
	`  anonymized_data = anonymization()`
	`  static_data,timeseries,semantic_links=metadata_decomposition(anonymized_data)`
	`  store_to_data_enclaves(static_data, timeseries, semantic_links)`

Outputs	Execute Eros4NRG Data Pipelines as job/containers

3.2 Collaborative Data Enclaves APIs and Plugins

Table 11: API Endpoint - List Static Information

List static information	
Description	Returns the static information (EMOTION EVs details, ASM Buildings and PVs details) stored in Graph Enclave
HTTP Method	GET
Endpoint	`/topology/source/${source_name}`
Parameters	Path parameter: select value for source_name to retrieve static information for either EMOTION or ASM Terni
Output	HTTP Response Code: 200

Table 12: API Endpoint - Get Dynamic Information for a specific EV

Get EVs values for a specific timeframe (EMOTION)	
Description	Returns the dynamic information (state of charge, velocity, charges_count, kms travelled, total amount of energy) generated by a specific EV
HTTP Method	GET
Endpoint	`/topology/source/emotion/ev/${ev_id}/dynamic/?start_date=${start_date}&end_date=${end_date}`

Parameters	ev_id: electric vehicle ID; start_date: datetime; end_date: datetime
Output	HTTP Response Code: 200

Table 13: API Endpoint - Get Building Dynamic Information

Get dynamic information for a specific building in a specific timeframe (ASM)	
Description	Returns the dynamic information (building usage hours, kWh consumed) generated by a specific Building
HTTP Method	GET
Endpoint	`/topology/source/asm/building/`**`${building_id}/`**`dynamic/?start_date=`**`${start_date}`**`&end_date=`**`${end_date}`**
Parameters	building_id: building ID; start_date: datetime; end_date: datetime
Output	HTTP Response Code: 200

Table 14: API Endpoint - Get PV Dynamic Information

Get dynamic information for a specific PV in a specific timeframe (ASM)	
Description	Returns the dynamic information (produced energy) related to a specific PV
HTTP Method	GET
Endpoint	`/topology/source/asm/pv/`**`${pv_id}`**`/dynamic/?start_date=`**`${start_date}`**`&end_date=`**`${end_date}`**
Parameters	pv_id: PV ID; start_date: datetime; end_date: datetime
Output	HTTP Response Code: 200

3.3 One-Stop-Shop (OSS) for Energy APIs, Plugins and User Interfaces

Table 15: User Interface - Login to Eros4NRG OSS

OSS Login	
Description	Login to OSS
HTTP Method	POST
Endpoint	`/superset/login`

Parameters	{
	“username”: **${username}**,
	“password”: **${password}**
	}

Output	HTTP Response Code: 201

Table 16: User Interface - ASM Terni Visualizations & Reports

ASM Visualizations & Reports	
Description	Navigate to ASM Terni Visualizations, Reports
HTTP Method	GET
Endpoint	`/superset/dashboard/asm/`
Parameters	Headers: Bearer **${access_token}**
Output	HTTP Response Code: 200

Table 17: User Interface - EMOTION Visualizations & Reports

EMOTION Visualizations & Reports	
Description	Navigate to EMOTION Visualizations, Reports
HTTP Method	GET
Endpoint	`/superset/dashboard/emotion/`
Parameters	Headers: Bearer **${access_token}**
Output	HTTP Response Code: 200

Table 18: API Endpoint - List AI/ML Experiments

List AI/ML Experiments	
Description	Retrieve a list of AI/ML Experiments
HTTP Method	GET
Endpoint	/experiments
Parameters	Headers: Bearer **${access_token}**
Output	HTTP Response Code: 200

Table 19: API Endpoint - List trained models for a specific experiment

List trained models for a specific experiment	
Description	Retrieve the list of all models trained for a specific experiment
HTTP Method	GET
Endpoint	`/experiments/`**`${experiment_id}`**`/models`
Parameters	Headers: Bearer **${access_token}** Experiment_id: experiment identifier
Output	HTTP Response Code: 200

Table 20: API Endpoint - Returns Information for a specific model

List trained models for a specific experiment	
Description	Returns the training parameters (e.g. learning rate, number of estimators), training metrics (e.g. MAE, MSE, RMSE), tags/metadata (e.g. short model description) and related artifacts (model file, python environment, requirements)
HTTP Method	GET
Endpoint	`/experiments/`**`${experiment_id}`**`/models/` **`${model_id}`**`/info`
Parameters	Headers: Bearer **${access_token}** Experiment_id: experiment identifier Model_id: model identifier
Output	HTTP Response Code: 200

Table 21: API Endpoint - Building Consumption Forecast

Building Energy Consumption (ASM)	
Description	Returns a 3 hour ahead forecast for ASM building energy consumption
HTTP Method	GET
Endpoint	`/building/`**`${building_id}`**`/forecast`

Parameters	Headers: Bearer **${access_token}** building_id: Building identifier
Output	HTTP Response Code: 200

Table 22: API Endpoint - ASM PV Energy Production Forecast

PV Energy Production (ASM)	
Description	Returns a 3 hour ahead forecast for ASM PV energy production
HTTP Method	GET
Endpoint	`/pv/${pv_id}/forecast`
Parameters	Headers: Bearer **${access_token}** pv_id: PV identifier
Output	HTTP Response Code: 200

Table 23: API Endpoint - Get EV State of charge Forecast

EV State of Charge (EMOTION)	
Description	Returns a 1 hour ahead forecast for EMOTION EV State of Charge
HTTP Method	GET
Endpoint	`/ev/${ev_id}/forecast`
Parameters	Headers: Bearer **${access_token}** ev_id: EV identifier
Output	HTTP Response Code: 200

Table 24: User Interface - Predictions Dashboard

Predictions Dashboard	
Description	Dashboard which lists trained experiments, models and models information. It also visualizes the predictions for a selected time frame.
HTTP Method	GET
Endpoint	`/superset/predictions/${emotion | asm}`
Output	HTTP Response Code: 200

4 Economic/Business Impact Potential

Eros4NRG aspires to take the leap forward in the big data in EU energy sector and internationally by establishing point of cross-secotrial big data integration and collaboration, which will respect privacy and data anonymization and ensure security. Bringing together all stakeholders involved in the big data value chain[31], and putting particular focus on the promising energy sector and Eros4NRG will deliver a solution for big data and intelligence creation leveraging high-quality data[32].

An initial list of these target groups, which at the same time are key actors that will directly benefit from the project results and outcomes, is depicted in the following table (Table 25).

Table 25: Target Audiences for Eros4NRG

Target Group	Description	Interest in the project
Energy value chain stakeholders	Power network operators; DSOs; Energy Analysts; aggregators; energy suppliers; Electromobility Managers	Leveraging Eros4NRG results to redesign their business models; Exploitation of Eros4NRG open-source results; Inspiration for new ideas and applications.
IT Industry & SMEs	IT Companies; web enterpreuners; software engineers; Data Analysts; Machine Learning Engineers	Excploitation of Eros4NRG open-source results; Inspiration for new ideas and applications; Integration of project results in corporate digital platforms to increase commercial offer.
Researchers and Academia	Individuals engaged in research initiatives and/or working in research/academic institutes	Further advance the project's research; Extension / reuse of the project's innovative technologies to other application domains, inspiration for future research initiatives based on the project concept and results.
Industry Associations & Technology Clusters	EU initiatives and clusters (like BDVA, IDSA, AIOTI and FIWARE); research communities, associations, federations (like IMS, EFFRA, IFIP, IEEE, NEM, ECTP)	Inclusion of project's results to collaborative research activities (roadmpa, white papers, etc); Dissemination of project's results to their members
Policy Makers & Standardisation Organisations	Policy-makers at any level like EC Directorates and Units, Ministries and Governments, Regulatory Agencies, Standardisation Organisations (CEN, ISO, ETSI, etc.)	Evaluation of the project's Social-Technological-Economic-Environmental-Political (STEEP) aspects; Definition of future research and innovation directions based on project's acquired knowledge; Inputs for standardisation activities

[31] https://www.big-data-value.eu/

[32] Azuatalam, Donald, et al. "Energy management of small-scale PV-battery systems: A systematic review considering practical implementation, computational requirements, quality of input data and battery degradation." Renewable and Sustainable Energy Reviews 112 (2019): 555-570.

As a result of the project implementation 10 Results will be delivered and are presented in Table 26, detailing the potential end-users as well the desired TRL.

Table 26: Eros4NRG Results

	Eros4NRG Results (KER #N)		Target Groups	Start / End TRL
Zero Trust Data Services	KER #1	IoT Streams Capturing	Energy value chain stakeholders; IT Industry & SMEs; Researchers & Academia	3/5
	KER #2	Anomaly Detection in IoT Streams		3/5
	KER #3	Workflow Orchestrator		3/5
	KER #4	Eros4NRG Data Pipelines		3/5
Collaborative Data Enclaves	KER #5	Eros4NRG Data Lake	Energy value chain stakeholders; IT Industry & SMEs; Researchers & Academia; Industry Associations & Technology Clusters	3/5
	KER #6	Eros4NRG Graph Enclave		3/5
	KER #7	Eros4NRG Dynamic Enclave		3/5
	KER #8	Secure Data Catalogue		3/5
One-Stop-Shop for Energy	KER #9	Visualizations & Reports Engine	Energy value chain stakeholders; IT Industry & SMEs; Researchers & Academia; Industry Associations & Technology Clusters; Policy Makers & Standardisation Organisations	3/5
	KER #10	Eros4NRG AI/ML Driven Analytics		3/5

Impact #1: Continuous monitoring of energy performance using state-of-the art data driven solutions

Eros4NRG will provide practical data-driven models based on high-quality data leveraging IoT/Edge data sources (**KER #1**, **KER #2**) from 2 pilot sites (ASM Terni, EMOTION). The improved data converted into valuable information will be used for the generation of AI/ML models (**KER #10**), designed to help during the decision-making process in the 2 target sites. Moreover, the automated anomaly detection mechanisms (**KER #4**) will reduce the time and efforts in term of hours of maintenance for data errors and correction.

Significance: The use of AI/ML driven models trained based on high-quality data achieves 5% - 10% energy savings[33]; Reduce the time for data maintenance in the energy sector 50%[34].

Impact #2: Use of real-time and reliable data from multiple IoT/Edge data sources

Eros4NRG will provide methods and standardized guidelines for data collection and setting up of databases, to improve and validate collected data from energy IoT/Edge sources. This allows to expand the potential for using energy monitoring data and public data as a screening tool in planning and assessing the potential for future energy assets affordable for everyone. To achieve this goal, Eros4NRG will deliver a fully fledged data warehouse which leverages Data Models, Interoperability mechaniscms and semantic links to optimize energy data organizations (**KER #5**, **KER #6, KER #7**). The delivery of

[33] Farzaneh, Hooman, et al. "Artificial intelligence evolution in smart buildings for energy efficiency." Applied Sciences 11.2 (2021): 763.
[34] Koseleva, Natalija, and Guoda Ropaite. "Big data in building energy efficiency: understanding of big data and main challenges." Procedia Engineering 172 (2017): 544-549.

Secure Data Catalogue (**KER #8**) will share high-quality data with EU-initiatives (e.g. EU Data Spaces, NEMO project).

Significance: (i) Number of different data sources: **7**, (ii) Use of different data models: **>=3,** (iii) Use of different warehousing methods: **>=2.**

Impact #3: Development of accurate methods that facilitate collection of data from energy sources

Existing data collection methods in the energy assets and sources are mainly conducted manually or using wireless sensors and networks but also cost-effective and insert errors and uncertainties to collected data thus decreasing the quality of components and services using these data. Eros4NRG will bring a practical and operational approach that combined ETLs and data ingestion procedures, followed by data quality and upgrade methodologies to collect heterogeneous and multiple data from energy-related data sources (e.g. PVs, EVs, smart buildings). These datasets, static and dynamic data, will be combined in the Eros4NRG Collaborative Data Enclaves (**KER #5**, **KER #6**, **KER #7, KER #8**) to perform and implement more robut analytics that support decision-making and impact assessment over the full energy life cycle.

Significance: Implement methods for anomaly detection: >=3; Integrate >=3 ingestion protocols.

Impact #4: Provide a One-Stop-Shop for Energy to enhance decision making in the energy sector

Eros4NRG OSS (**KER #9, KER #10**) consists of two main modules:

- Visualizations & Reports Engine: This module provides visualizations and charts that depict the energy data landscape and the results of AI/ML model executions.
- AI/ML Driven Analytics: This module includes analytics tailored for the two target pilot sites (ASM Terni and EMOTION). It offers models for:
    - 3-hour ahead building energy consumption forecasting
    - 3-hour ahead PV energy production forecasting (ASM Terni)
    - 1-hour ahead EVs state of charge forecasting (EMOTION)

In the Eros4NRG One-Stop-Shop (OSS) for Energy, stakeholders can interact with Eros4NRG components, high-quality data, and reliable models. This interaction will enable end-users to make better informed and accurate decisions, reduce uncertainties with advanced AI/ML models, and retrieve early insights to optimize energy consumption in smart buildings and EVs' batteries state of charge and usage.

Significance: High-Quality AI-based services deployed in **2** pilot sites; Train **>= 3 models** for each pilot site; Increase stakeholder awareness using data-driven decision making at 50%[35].

[35] Ahmad, Tanveer, et al. "Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm." Renewable and Sustainable Energy Reviews 160 (2022): 112128.

5 Conclusions & Next Steps

This document outlines the initial work carried out in the first phase of the Eros4NRG project, specifically **WP1 – "DESIGN**" It details the comprehensive design specifications of Eros4NRG, including an in-depth presentation of Use Cases, Non-Functional and Functional Requirements, the resulting system architecture, and the elaboration of APIs, Plugins, and User Interfaces that will be part of the Eros4NRG platform. Additionally, the document discusses the Economic/Business Impact of Eros4NRG, highlighting its target audience, Key Exploitable Results (KERs), and how the project results will bring business and economic potential to the energy market.

This document will serve as the foundation for the next phase of the Eros4NRG project, **WP2 – "DEVELOP"** during which the core components of Eros4NRG will be developed following the defined specifications and system architecture. A more detailed view of the Eros4NRG system architecture will be provided in document **D2 – "Eros4NRG MVP pre-pilot release"**.